

UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
COORDENAÇÃO DE PESQUISA

Programa Institucional de Bolsas de Iniciação Científica - PIBIC
POSGRAP/PROEST/UFS

PROJETO PVE4749-2016 - IHCasa - Modelo de Interação Humano-Casa
baseado em Inteligência Artificial e Processamentos de Imagens para a
Detecção, Identificação e Interação com Funções Residenciais Simples

PLANO DE TRABALHO: Levantamento e Reconhecimento de Funções
Básicas para Interação Residencial

Área de Concentração: Computação

Bolsista: Maurício José Miranda Guimarães
Nº Matrícula: 201320019574

Orientador: Prof. Dr. Alcides Xavier Benicasa
DSI - Departamento de Sistemas de Informação

Relatório Final PIBIC 2016-2017
Período AGO/2016 à JUL/2017

Sumário

1	Introdução	1
2	Revisão da Literatura	4
2.1	Interação Natural	4
2.2	Processamento de Imagens	5
2.3	Linguagem de Programação Utilizada e Captura de Imagens	5
2.4	Automação Residencial	6
2.4.1	Arduino	7
2.4.2	ESP8266	8
2.5	Trabalhos Relacionados	9
3	Modelo Proposto	12
3.1	Estrutura Geral do Modelo	12
3.1.1	Modelo de Visão Computacional	13
3.1.2	Câmera	17
3.1.3	Seleção de Ícones	18
3.1.4	Plataforma de comunicação	18
3.1.5	Reconhecimento de Funções Básicas para Interação Residencial . . .	21
3.1.6	Reconhecimento de Tentativas de Interação	24
4	Resultados	26
5	Conclusões	30

Capítulo 1

Introdução

A necessidade de proporcionar conforto e comodidade, até mesmo em situações do cotidiano é notável e tem se intensificado com o passar dos anos. Este fator, juntamente com os constantes desenvolvimentos em tecnologia, o rápido avanço na miniaturização de dispositivos, o surgimento de ferramentas para a comunicação sem fio e processadores portáteis, impulsionam o desenvolvimento da automação residencial. O objetivo principal é facilitar tarefas diárias como, por exemplo, acender uma lâmpada, ligar um ventilador, ligar uma televisão ou até mesmo um ar-condicionado. Segundo Chan et al. (1995), Lee and Choi (2003), Muratori and Dal Bó (2011) e Ahmim et al. (2016), automação residencial consiste em um conjunto de sensores, equipamentos, serviços e diversos sistemas tecnológicos integrados com intuito de auxiliar às necessidades básicas de segurança, comunicação, gestão energética e conforto de uma habitação. Segundo Ricquebourg et al. (2006), uma casa inteligente é entendida como um sistema de integração, que tira proveito de uma variedade de técnicas, tais como computadores, e algum tipo de comunicação com intuito de conectar todos os subsistemas internos que são conectados a eletrodomésticos e aparelhos elétricos domésticos como um todo.

Atrelado ao notável crescimento do interesse com relação à automação residencial, há também uma variedade de alternativas, no que se trata de comunicação com os aparelhos residenciais. Tendo como exemplo desde soluções proprietárias, como pode ser visto em Creston (2016) e Lutron (2016), circuitos construídos utilizando *Raspberry Pi* (Celebre et al., 2015), bem como estruturas com Arduino (Silva, 2014), (Mittal et al., 2015), (Morshed et al., 2015), (Mittal et al., 2015) e (Kumar and Solanki, 2016). Um outro ponto que merece ser destaque é a variedade no que se refere à forma de conexão aplicada entre os dispositivos. Observa-se trabalhos que utilizam comunicação através da tecnologia *bluetooth*, a partir transmissores e receptores infravermelhos, ou mesmo através de redes sem fio.

Com o objetivo de tornar a interação homem e residência mais simples e intuitiva, diversas pesquisas têm apresentado novas funcionalidades de interatividade, sendo baseadas em comandos de voz, em expressões faciais, bem como através da captura e interpretação de gestos. Dentre estas, a última citada é o principal ponto de investigação dessa pesquisa.

De acordo com Correia (2014), desde a década de 80 a utilização de interação gestual é apresentada na literatura, sendo esta uma forma de interação flexível, realista e natural. Segundo Valli (2007), sistemas que respondem a gestos livres entendem ações naturalmente utilizadas pelas pessoas para se comunicar, permitindo aos usuários interagir facilmente com o ambiente a seu redor. Com isso, o uso de gestos para a comunicação com dispositivos diminui a carga cognitiva, resultando em um modo bastante natural e intuitivo de interação.

Segundo Aggarwal and Cai (1997), a área de Interação Homem-Computador (IHC) percebe a relevância da comunicação corporal desde suas primeiras investigações, realizando desde o início de 1980 estudos a respeito de questões quanto ao rastreamento do corpo humano. O uso de sensores, apesar de facilitar a identificação de gestos, acaba acarretando maior custo na solução desenvolvida e ainda, de acordo com Blake (2012), a necessidade da utilização de algum dispositivo vai contra o conceito de liberdade em interação natural.

Sendo assim, este projeto visa desenvolver um modelo que pode ser considerado como uma variação da IHC (Interação Humano-Computador), denominado aqui por Interação Humano-Casa (IHCasa). É também foco deste trabalho implementar um modelo inicial de inteligência, o qual deve ser intuitivo, de baixo custo e eficiente, capaz de permitir a interação com funções residenciais simples.

De uma maneira geral, considera-se como uma função interativa residencial simples, por exemplo, a alteração do estado de uma lâmpada, que pode estar ligada ou desligada. Para a alteração de estado, ou seja, para que seja possível a interação entre o usuário e o estado da lâmpada, faz-se necessária a existência de diversos dispositivos prévios, sendo estes: interruptores, fiações, conduítes, cortes na estrutura da parede, etc. Neste trabalho, é proposto a exclusão de todos estes artefatos, de modo que o local de interação seja representado por uma imagem ou símbolo representativo, referente a uma função específica.

Considerado como uma pesquisa prévia a este projeto, em Santos et al. (2016) foi publicado por nosso grupo de pesquisa no XIII Encontro Nacional de Inteligência Artificial e Computacional (ENIAC'2016), um modelo de interação visual com imagens a partir de técnicas de processamento de imagem e inteligência artificial para identificar, classificar e permitir a interação com objetos de maneira natural, isto é, sem o uso de sensores ou aparelhos especializados. O modelo desenvolvido já fora objeto de pesquisa e desenvolvimento anteriormente com base em uma ambiente estático, tendo sido ampliado e incorporado para uma aplicação mais dinâmica e flexível.

O atual projeto de pesquisa visa prover uma abordagem de interação humano-casa (IHCasa) com base nos princípios de interação dinâmica e por meio de gestos, baseada em visão computacional e dispensando a utilização de sensores e dispositivos especiais. Para tanto, utilizaremos inicialmente o modelo já em desenvolvimento por nosso grupo de pesquisa no que refere-se à visão computacional para o novo modelo aqui proposto.

Contudo, o modelo foi adaptado de maneira a lidar com novos ambientes e configurações, menos controlados e mais complexos, devendo funcionar tanto diurno como noturnamente, ampliando ainda mais o seu escopo operacional. Adicionalmente, a aplicação comunica-se com uma placa controladora, critério que irá tratar da automação presente neste projeto, visando enviar instruções necessárias à execução de comandos aos aparelhos eletrônicos e demais dispositivos considerados neste trabalho a exemplo de uma casa inteligente. O modelo atual foi publicado na 19ª Conferência Internacional sobre Interação Homem-Computador (HCI Internacional 2017) realizada em Vancouver, Canadá, no período de 9 a 14 de julho de 2017, com o título de Human-House Interaction Model Based on Artificial Intelligence for Residential Functions (Santos et al., 2017). Tal publicação consta como um capítulo do livro HCI International 2017 - Posters' Extended Abstracts.

O projeto foi dividido em três planos de trabalho. No primeiro plano, os objetivos específicos estão relacionados com o estudo e seleção de mecanismos para a captura de imagens através de câmera de vídeo em tempo real, tendo ainda neste sido abordado o estudo sobre algoritmos de processamento de imagens para o tratamento e obtenção de informações relevantes da cena que possam identificar regiões candidatas a alvos (*targets*), podendo ser considerado como mecanismo de atenção visual. No segundo plano, os objetivos específicos foram relacionados ao estudo e aplicação de métodos avançados de segmentação de imagens. No terceiro e último plano, o qual se refere o presente relatório, os objetivos específicos foram relacionados ao estudo e aplicação de métodos tradicionais de treinamento de padrões. Ainda neste plano, abordou-se o estudo de funções de interação residencial existente na forma manual necessária à interação do usuário com a residência.

O seguinte relatório encontra-se organizado da seguinte forma: Inicialmente, realizou-se a revisão da literatura necessária para o embasamento teórico utilizado durante o desenvolvimento desta proposta, com o estudo sobre abordagens de automação residencial. Na revisão da literatura também é apresentada a linguagem de programação utilizada e o funcionamento da placa computacional *Arduino*, adotado neste projeto como placa controladora. O modelo de visão computacional e os resultados obtidos neste plano de trabalho são abordados no capítulo 3. Os resultados gerais do projeto de pesquisa são apresentados e analisados no Capítulo 4. O Capítulo 5 conclui o trabalho apresentando um resumo dos principais resultados obtidos e melhorias a serem executadas.

Capítulo 2

Revisão da Literatura

Neste capítulo serão apresentados os embasamentos teóricos que foram utilizados para o desenvolvimento deste projeto.

2.1 Interação Natural

Os grandes avanços tecnológicos ocorridos principalmente nas últimas décadas possibilitaram a concepção de uma nova modalidade de interface, disponível em aplicações como telas sensíveis ao toque, reconhecimento de comandos de voz, reconhecimento de gestos, entre outras. Este é conhecido como Interfaces Naturais de Usuário (NUI). De acordo com Blake (2012), NUIs podem ser definidas como “interfaces projetadas para reusar habilidades existentes para a interação direta com o conteúdo”.

A definição proposta por Blake (2012) evidencia três importantes conceitos a respeito de NUI. Primeiro, NUIs são projetadas, ou seja, requerem que sejam premeditadas e que sejam realizados esforços prévios para sua concepção. É preciso assegurar que as interações em uma NUI sejam apropriadas tanto para o usuário quanto para o conteúdo e o contexto. Segundo, NUIs reutilizam habilidades existentes. Durante vários anos, os usuários têm praticado a comunicação, verbal ou não verbal, além de interações com o ambiente. O poder computacional e tecnologia evoluíram ao ponto em que é possível tirar vantagem dessas habilidades. NUIs fazem isso ao permitir interações por meio de ações intuitivas como tocar, gesticular e falar. Finalmente, NUIs têm interação direta com o conteúdo, ou seja, o foco da interação está no conteúdo e na interação direta com ele. Isso não significa que a interface não possa ter controles, como botões ou caixas de seleção, se necessário. Significa apenas que esses controles devem ser considerados secundários, comparados ao conteúdo, e que a direta manipulação do conteúdo deve ser o método de interação primário.

Dessa forma, como afirma et al. Dias (2013), uma interface NUI exige apenas que o usuário seja capaz de interagir com o ambiente por meio de interações previamente já conhecidas pelo mesmo. Este tipo de interface exige aprendizagem, porém esta é facilitada, pois não exige que o usuário seja apresentado a um novo dispositivo.

2.2 Processamento de Imagens

O uso de câmeras para detecção da interação do usuário com uma aplicação tem sido muito comum em sistemas interativos, em especial nas aplicações baseadas em sensoriamento sem toque, onde, além do uso da câmera, algoritmos de processamento de imagens são utilizadas para o tratamento e obtenção de informações relevantes da cena.

A área de processamento de imagens digitais tem atraído grande interesse nas últimas décadas. A evolução da tecnologia digital aliada ao desenvolvimento de novos algoritmos, capazes de processar sinais bidimensionais, vem permitindo uma gama de aplicações cada vez maior (Morales et al., 2003), como por exemplo, na medicina principalmente na ajuda de diagnósticos e em cirurgia guiada por computador, em geoprocessamento, radares de trânsito, sensoriamento remoto na visualização do clima de uma determinada região, na arquitetura e nas engenharias (elétrica, civil, mecânica), entre outros (Morgan, 2008).

Para e Woods R. E. Gonzalez (2010), o interesse em métodos de processamento de imagens digitais decorre de duas áreas principais de aplicação: melhoria da informação visual para a interpretação humana e processamento de dados para percepção automática através de máquinas. Segundo Grand (2005), na abordagem de Gonzáles e Woods, a primeira categoria concentra-se em técnicas para melhora de contraste, realce e restauração de imagens danificadas. A segunda categoria concentra-se em procedimentos para extrair de uma imagem informação de forma adequada, para o posterior processamento computacional. É na segunda categoria, ou seja, na percepção automática por máquinas, que se enquadra o trabalho aqui descrito.

2.3 Linguagem de Programação Utilizada e Captura de Imagens

Para implementar a aplicação foi utilizada a linguagem Java, que é uma linguagem de programação e plataforma computacional lançada pela primeira vez pela Sun Microsystems em 1995 (Java, 2014). A linguagem Java segue o paradigma orientado a objetos, isto é, de acordo com Javafree (2014), um sistema construído em Java é composto por um conjunto de classes e objetos bem definidos que interagem entre si, de modo a gerar o resultado esperado.

Cada classe é a estrutura de uma variável, ou seja, um tipo de dado. Nela, são declarados atributos e métodos que poderão ser executados ou acessados nas instâncias da mesma classe. As classes possuem uma função muito importante na modelagem orientada a objetos, elas dividem o problema, modularizam a aplicação e baixam o nível de acoplamento do software. Variáveis de uma classe são chamadas de instâncias de classe, ou objetos. Um Objeto, ou instância, é uma entidade cujas informações podem incluir desde suas características até suas atividades, ou seja, uma abstração de um domínio de um problema (Javafree, 2014).

A justificativa para a escolha da tal linguagem de programação se dá devido à vasta documentação disponível na internet, a disponibilidade de IDEs poderosas e gratuitas, bem como diversas bibliotecas que podem ser necessárias para o funcionamento da aplicação.

Para a captura de imagens foi utilizada a biblioteca WEBCAM-CAPTURE de Bartosz (2013), que permite o acesso a câmeras de vídeo diretamente a partir do código Java, possibilitando a leitura de imagens e detecção de movimentos, oferecendo suporte à diversas plataformas.

Para a programação da placa Arduino foi utilizado o software de programação Arduino IDE, disponível para os sistemas operacionais Windows, Linux e Mac OS. A linguagem de programação utilizada no desenvolvimento para o Arduino tem como base a linguagem C. A interface desta IDE se destaca por ser rápida, de simples utilização e intuitiva.

2.4 Automação Residencial

A automação residencial, também conhecida como domótica, tem como objetivo de facilitar atividades no ambiente doméstico. Esta tem crescido bastante, sempre visando buscar soluções que proporcionem mais conforto e comodidade para os moradores de uma residência.

A automatização de dispositivos residenciais fornece uma série de benefícios aos usuários tais como a facilidade de comunicação, segurança, entretenimento e, principalmente, conforto e economia de energia. Com a tecnologia mais inserida no cotidiano das pessoas de todas as faixas etárias, a procura por residências automatizadas tem se intensificado.

No Brasil, o cenário da automação residencial vem ganhando espaço nos últimos anos, e, de acordo com TechinBrazil (2015), representa um mercado de consumo que adquire lucros próximos a 500 milhões de reais por ano, além de um crescimento anual de 30%.

Desta forma, o conceito de automação residencial baseia-se em instalações onde existe a integração de sistemas residenciais, bem como a capacidade de execução de funções através de instruções programáveis. De acordo com Sombra (2016), os sistemas de uma residência suscetíveis à automação podem ser:

- Instalação Elétrica, que corresponde: iluminação, persianas e cortinas e gestão de energia;
- Sistema de segurança: alarmes contra intrusos, alarmes técnicos (fumaça, vazamento de gás, inundação), câmeras de segurança e controle de acesso;
- Sistemas de entretenimento: áudio e vídeo, som ambientes, imagens e sons sob demanda;
- Utilidades: irrigação, ar condicionado, aspiração central, aquecimento de água, bombas;
- Sistemas de telefonia.

Conclui-se, portanto, que a automação residencial consiste no controle de um ou mais itens acima citados através de um computador ou algum microcontrolador. Ao passo que, tais dispositivos devem ser monitoráveis e controláveis remotamente, a partir da utilização de recursos de comunicação.

A seguir, os conceitos pertinentes acerca das tecnologias utilizadas no modelo proposto serão descritos.

2.4.1 Arduino

O Arduino é uma placa controladora de simples utilização e com alta aplicabilidade. Uma característica de grande importância é que consiste em plataforma livre (*open-source*), fator que implica em um *hardware* com baixo custo, além de possibilitar que o desenvolvimento de *softwares* seja gratuito (Durfee, 2011). O seu ambiente de programação é simples e amigável, e pode ser encontrado para os sistemas Windows, Mac e Linux.

As placas Arduino são capazes de ler sinais de entrada, tais como, sensores, botões, ou até mesmo comandos através da internet e transformá-los em uma saída, como por exemplo, ativar um motor, ligar uma luz, ou acionar algum dispositivo. Podemos concluir que o Arduino consiste em um plataforma extremamente versátil, oferecendo uma gama de possibilidades e alternativas.



Figura 2.1: Arduino Uno (Durfee, 2011)

Existem diversos tipos de modelos de Arduino disponíveis no mercado, estes se diferenciam, principalmente, quanto à robustez do sistema e quantidade de portas de entrada e saída. Os principais modelos são:

- Arduino Uno
- Arduino Leonardo
- Arduino LilyPad
- Arduino Mega
- Arduino Nano
- Arduino Mini

- Arduino Mini Pro.

A escolha pela utilização do Arduino se deu pela sua vasta utilização na literatura, como pode ser visto em Solanki and Desai (2011), Silva (2014), Sumanta Dey and Dutta (2015), Morshed et al. (2015), Mittal et al. (2015), Kumar and Solanki (2016) e Deshmukh et al. (2016). Também foi levado em consideração o seu baixo custo, quando comparado a plataformas com o mesmo propósito, a sua versatilidade, e, principalmente, pela facilidade de desenvolvimento e construção de programas. A opção pelo uso do modelo Uno, levou em consideração os fatores investimento e demanda, visto que tal modelo atende as necessidades e requisitos do projeto, possuindo o menor valor comercial.

2.4.2 ESP8266

Considerando um dos objetivos, que consiste na redução na quantidade de cabos, fiações e conexões, procurou-se uma solução que permitisse conectividade sem fio. Dentre os diversos módulos existentes no mercado e na literatura, destaca-se o módulo Wi-Fi ESP8266, exibido no Figura 2.2.

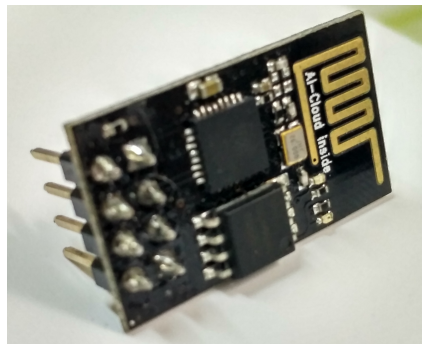


Figura 2.2: Módulo ESP8266

Inicialmente, dois aspectos deste módulo se destacam, são eles, o seu tamanho reduzido e o seu baixo custo, possuindo uma faixa de preço de cinco dólares Mehta (2015), a Tabela 2.1 exibe um comparativo entre os preços de diversos modelos existentes no mercado em relação ao ESP8266. Além disso, um ponto de extrema importância é a sua facilidade de integração com diversas plataformas, através da comunicação serial, fato que possibilita a conexão e compatibilidade com o Arduino, como pode ser observado em Baig et al. (2016) e Delazeri et al. (2017).

O ESP8266 possui uma vasta gama de recursos, podendo ser considerado como uma plataforma robusta, projetada com o intuito de utilizar uma quantidade mínima de componentes. O módulo se comunica através de conexões TCP e UDP, permitindo múltiplas conexões simultâneas.

Na seção a seguir serão expostos os trabalhos relacionados para estudo e comparação.

Tabela 2.1: Comparativo de valores entre diversos módulos. Traduzido de Mehta (2015)

Módulo Ethernet para Arduino	\$ 60
ZigBee	\$ 25
Módulo Wi-Fi Sparkfun	\$ 40
Módulo Wi-Fi para Arduino	\$ 80
Módulo Wi-Fi Huzzah	\$ 40
ESP8266-12	\$ 7
ESP8266-01	\$ 5

2.5 Trabalhos Relacionados

Inicialmente, dentre os mais diversos tipos de alternativas e soluções, serão apresentados estudos que se propõem a automatizar uma residência, porém, sem a utilização de qualquer tipo de interpretação de gestos ou processamento de imagem. A Tabela 2.2 apresenta trabalhos que possuem essas características.

Tabela 2.2: Exemplos de trabalhos que não aplicam métodos de interpretação de gestos ou processamento de imagem no processo de automação residencial

Título do trabalho	Autor(es)	Ano
Sistema de automação residencial de baixo custo	Silva	2014
Microcontroller based home automation system using Bluetooth, GSM, Wi-Fi and DTMF	Morshed et al.	2015
A voice-controlled multi-functional Smart Home Automation System	Mittal et al.	2015
IP network based home automation system using AVR microcontroller	Ahmed et al.	2016
Voice and Touch Control Home Automation	Kumar e Solanki	2016

No ano de 2014, Silva propôs um sistema de baixo custo, composto por sensores controlados por meio de um aplicativo móvel para Android. A arquitetura é composta por um Arduino, além da utilização de protocolos Wi-Fi para a transmissão de informações. Em 2015, Morshed et al. propuseram um modelo de automação residencial que também utiliza Arduino, capaz de responder a comandos enviados através de *Bluetooth*, internet e mensagens de texto.

Observa-se em Ahmed et al. (2016) a proposta de um modelo centralizado, que possui funcionamento através de cabo *Ethernet* ou Wi-Fi, e a arquitetura é composta por três camadas. A camada de entrada é responsável por receber comandos do usuário através de um computador ou dispositivo móvel e enviá-los para a camada de controle. A camada de controle tem o papel de interpretar o comando solicitado e enviar para a camada atuadora, esta última aciona os dispositivos da residência.

Os trabalhos propostos por Mittal et al. (2015) e Kumar and Solanki (2016) possuem características em comum e merecem receber destaque, pois apresentam uma forma de interação natural. Os sistemas são capazes de reconhecer comandos de voz dos usuários, dessa forma, sendo possível controlar as aplicações domésticas. Um outro ponto impor-

tante presente em ambas pesquisas é a utilização do Arduino Uno.

A utilização de gestos mostra-se como uma maneira intuitiva de comunicação e tem sido utilizada em diversas pesquisas relacionadas com o desenvolvimento de modelos de interação para automação residencial, proporcionando o uso próximo da interação natural. Dessa forma, não se faz necessário um treinamento prévio por parte do usuário para a utilização do dispositivo, pois seus comandos são intuitivos. A Tabela 2.3 exibe trabalhos que possuem como objetivo a interpretação de gestos para a realização de ações em aplicações domésticas, além de destacar as principais características destes. Na última linha da tabela, encontra-se as características do trabalho aqui proposto.

Tabela 2.3: Comparativo de trabalhos que interpretam gestos durante o processo de automação residencial

Título do trabalho	Autor(es)	Ano	Ambiente	Sensor especial	Câmera
Hand Gesture Based Remote Control For Home	Solanki e Densai	2011	Claro e escuro	Não	Sim
Gesture Controlled Home Automation For Differently Challenged People	Prabhuraj e Saravanakumar	2014	Claro e escuro	Sim	Não
To Analyze Hand Gesture Recognition For Wirelessly Electronic Device Control	Pannase e Ugale	2014	Claro	Não	Sim
Gesture Controlled Home Automation	Sumanta Dey et al.	2015	Claro	Não	Sim
Automatic Hand Gesture Based Remote	Bharambe et al.	2015	Claro e escuro	Sim	Não
Automatic Hand Gesture Based Remote	Deshmukh et al.	2016	Claro e escuro	Sim	Sim
IHCasa - Modelo de Interação Humano-Casa baseado em Inteligência Artificial e Processamento de Imagens	Santos	2017	Claro e escuro	Não	Sim

A captura dos gestos pode ser realizada através da utilização de dispositivos especiais, tais como, sensores, acelerômetros e luvas. Tais acessórios facilitam a interpretação de gestos. Porém, como afirma Erol et al. (2005), aumentam bastante o custo do sistema. Nesse sentido, destaca-se o trabalho desenvolvido por Prabhuraj and Saravanakumar (2014), que utiliza uma luva composta por vários sensores e acelerômetros para capturar gestos e responde a gestos presentes na Língua Americana de Sinais. Os dados capturados são classificados através da utilização de redes neurais artificiais.

No ano de 2015, seguindo a linha de utilização de sensores especiais, observa-se em Apoorva Bharambe and Vitekar (2015) um sistema composto por um acelerômetro para capturar os gestos, um microcontrolador responsável por identificar as informações coletadas, um transmissor e um receptor infra-vermelho, este último tem o papel de acionar o aparelho desejado. Vale ressaltar que existe uma limitação quanto à distância, além do fato de que o transmissor e o receptor infravermelho devem estar totalmente alinhados.

Deshmukh et al. (2016) apresentaram um modelo de processamento de imagem em tempo real para reconhecimento de gestos pertencentes à Língua Americana de Sinais. A sua arquitetura é composta por uma *webcam* modificada para possibilitar a visualização de luzes infravermelho, um acelerômetro para auxiliar a identificação dos gestos, transmissor e receptor infravermelho, além de um Arduino Uno. Aqui também observa-se a limitação quanto a distância e alinhamento entre o transmissor e receptor infravermelho.

Em contrapartida à utilização de acessórios especiais, os métodos baseados em Visão Computacional necessitam apenas de uma câmera de computador para a captura de ima-

gens e tem como objetivo a interpretação automática de cenas complexas Jain and Dorai (1997); Raut et al. (2012). Sendo assim, em 2011, Solanki and Desai (2011) propuseram um sistema com o objetivo de realizar funções presentes em controles remoto de uma televisão através de gestos realizados com a mão. O processamento da imagem e interpretação dos gestos ocorrem em tempo real, onde as imagens são capturadas por meio de uma *webcam*, e também é utilizada uma iluminação infra-vermelho. A etapa de interpretação das imagens capturadas é realizada em um *notebook*, posteriormente os dados são enviados para um Arduino, sendo este último responsável por enviar os comandos para o aparelho desejado através de um módulo transmissor infravermelho. Essa proposta é limitada a funções presentes em uma televisão, além de utilizar sinais de infravermelho para a transmissão de informação, como já citado acima, um meio de transmissão limitado quanto a distância e alinhamento dos dispositivos.

Unindo os conceitos de Visão Computacional e aplicações móveis, no ano de 2015, Sumanta Dey and Dutta (2015) propuseram um trabalho onde a captura e interpretação de gestos é realizada utilizando um dispositivo móvel. É gerado um sinal binário a partir da imagem capturada pelo celular, este é classificado e enviado utilizando *Bluetooth* para um Arduino, que por sua vez é responsável por acionar os aparelhos da residência.

Ainda se tratando de métodos baseados em Visão Computacional, em 2014, Pannase and Ugale (2014) desenvolveram uma pesquisa, a qual o objetivo foi detectar rapidamente gestos realizados pela mão. Para isso, eles utilizaram algoritmos para segmentar a região central da mão, detecção do posicionamento dos dedos, e, por fim, a classificação do gesto através de rede neural. A transmissão dos comandos ocorre através de rede sem fio. Outro aspecto deste estudo é a quantidade limitada de gestos, bem como a mão deve ser posicionada exatamente a frente da câmera.

Este trabalho propõem um modelo de interação entre um humano e um ambiente residencial, dispensando a utilização de sensores ou dispositivos especiais, proporcionando uma solução mais natural e intuitiva para controlar funções de uma casa, por exemplo, desligar e ligar um televisão, lâmpadas, abrir e fechar portas e janelas, etc. O sistema de interação deve ser capaz de operar em ambientes claros e escuros. Para isso, uma *webcam* foi adaptada para capturar luz infravermelha. Uma outra característica importante é a utilização de uma rede Wi-Fi para a transmissão de informações.

O próximo capítulo apresentará alguns detalhes do funcionamento do modelo, além de expor as suas principais características.

Capítulo 3

Modelo Proposto

Neste capítulo, aspectos gerais desenvolvidos durante o projeto serão expostos. O modelo aqui proposto tem por objetivo prover uma solução inicial de interação homem-casa por meio de gestos, sem a utilização de sensores especiais ou outros dispositivos, oferecendo um ambiente de interação natural para o controle de funções de uma residência inteligente, possibilitado pelo uso de técnicas de inteligência artificial de visão natural e interpretação de cenas complexas obtidas por meio de capturas de imagens. A análise da cena busca identificar ações de interação executadas por usuários aos alvos presentes no ambiente observado.

3.1 Estrutura Geral do Modelo

Conforme mencionado, o modelo aqui proposto tem como objetivo controlar as funcionalidades de uma residência, como ligar e desligar aparelhos eletrônicos, luzes, abrir e fechar portas, entre outras. Esta interação deve acontecer através das opções de interação distribuídas pelo ambiente de maneira desejada pelo usuário.

Como pode ser observado na Figura 3.1, inicialmente, no ambiente encontram-se os alvos de interação, que é um conjunto de objetos alvos, os quais são distribuídos pelo ambiente de maneira desejada pelo usuário e possuem uma representação sugestiva quanto ao dispositivo ou função que acionam. A câmera, por sua vez, deve estar direcionada para a região do ambiente onde se encontram os alvos de interação. No modelo desenvolvido, utiliza-se uma câmera *webcam* comum, adaptada para poder capturar também com o auxílio de luz infravermelha. Em seguida, representado na arquitetura como Computador Central, encontra-se a aplicação principal, responsável por receber os dados de imagens capturadas pela câmera de vídeo e realizar os processos de Visão Computacional e identificação da interação acionada, enviando, ao final, o comando para a placa controladora. A placa controladora conecta-se tanto ao computador central, como aos dispositivos passíveis de controle e interatividade na casa inteligente, sendo responsável pela recepção do comando identificado pela aplicação no computador central e posterior disparo da ação ao seu respectivo dispositivo desejado. Como dito anteriormente, a placa

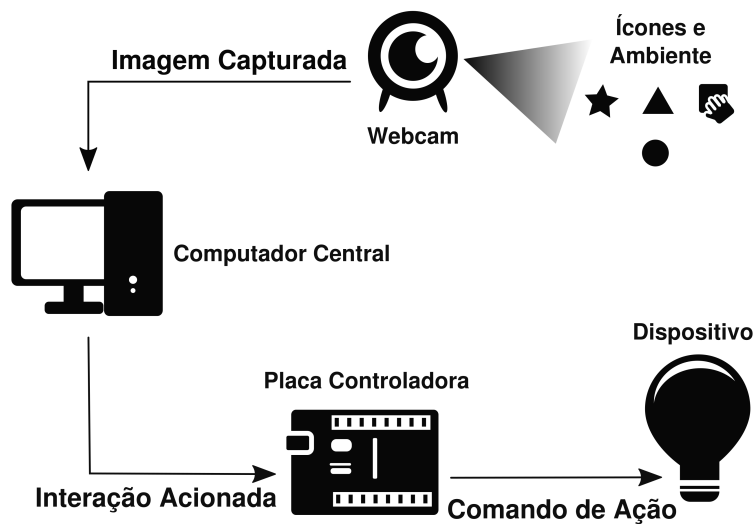


Figura 3.1: Fluxo de funcionamento do modelo proposto

controladora utilizada durante o projeto foi o Arduino Uno.

Adicionalmente, a aplicação principal se comunica com uma câmera de vídeo que, ao receber uma imagem da câmera a processa, retornando como resultado a identificação de uma tentativa de interação e sua ação específica, enviando esta informação para a placa controladora que, por fim, executa a ação junto ao respectivo aparelho ou função desejada como, por exemplo, ligar ou desligar uma lâmpada.

O funcionamento do modelo pode ser observado na Figura 3.2, onde foi utilizada uma maquete com o intuito de ilustrar uma residência. É possível observar o acionamento de uma lâmpada ao passo que ocorre uma interação com um objeto de interação.



Figura 3.2: Ilustração do funcionamento do modelo utilizando uma maquete simulando uma residência

3.1.1 Modelo de Visão Computacional

A aplicação no computador principal incorpora o modelo desenvolvido por nosso grupo de pesquisa, publicado em Santos et al. (2016). Este modelo incorpora diferentes técnicas de inteligência artificial, visão computacional e processamento de imagem para a identificação e classificação dos objetos específicos (neste caso, os ícones de interação) em um ambiente complexo e não controlado.

As etapas do modelo de visão consistem em pré-processamento inicial, responsável por destacar os *pixels* de cor vermelha, em seguida, é aplicado o método de atenção visual para identificar regiões de interesse presentes na imagem. Neste caso, é utilizada uma modificação do modelo de Itti et al. (1998), onde este foi simplificado para responder apenas ao estímulo de cor. As áreas de interesse são marcadas como sementes, que são entradas para o processo de segmentação. Uma vez que a segmentação obtém os elementos individuais de cada ícone, estes são classificados através de redes neurais artificiais, com intuito de atribuir uma ação de interação aos ícones. O que segue é a rotina de interação, esta continuamente captura imagens buscando possíveis ações de interação do usuário. As interações são reconhecidas como mudanças no coeficiente de dissimilaridade entre os histogramas das regiões onde os objetos de interação foram identificados.

O funcionamento do modelo de visão computacional é apresentado na Figura 3.3.

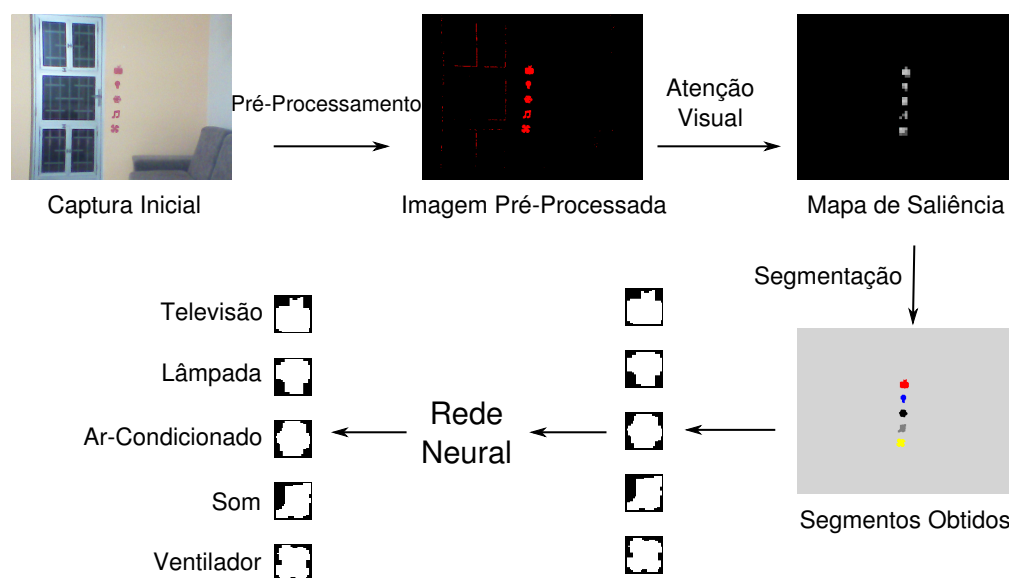


Figura 3.3: Esquema de funcionamento do modelo de visão computacional

Pré-Processamento

Como fase de pré-processamento foi utilizada uma abordagem para a saturação dos *pixels* com coloração vermelha. Esta etapa possui intenção de eliminar possíveis ruídos na imagem capturada, bem como possibilitar uma melhoria no processo de atenção visual, direcionando-o apenas para os objetos que forem destacados.

Atenção Visual

A atenção é um mecanismo de seleção cognitiva que consiste em selecionar o que possui importância para o organismo a partir de uma entrada composta por uma pluralidade de informações. Humanos e animais enfrentam constantemente a tarefa de processar uma grande quantidade de informações através do sistema visual. A quantidade de informação, por sua vez, é demasiadamente grande para que todos os detalhes possam ser processados,

fazendo-se necessário que haja uma filtragem. Segundo Benicasa (2013), pode-se afirmar que o processo de seleção visual seleciona apenas um subconjunto da informação sensorial disponível, na forma de uma região circular do campo visual, conhecida como foco de atenção.

O estudo de mecanismos computacionais capazes de determinar regiões relevantes em ambientes repletos de informações visuais tem crescido bastante, principalmente nas últimas três décadas. Dentre as teorias que influenciam estas, merecem destaque a *Feature Integration Theory* Treisman and Gelade (1980) e *Guided Search Model* Wolfe (1994).

Os modelos computacionais de atenção visual possuem estrutura similar, a mesma está ilustrada na Figura 3.4. A ideia principal é calcular um conjunto de características em paralelo e então fundi-las em uma representação comumente chamada de mapa de saliência Frintrop et al. (2010).

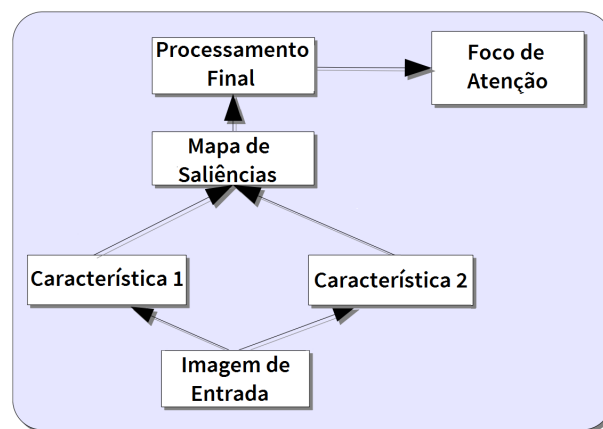


Figura 3.4: Estrutura genérica de um modelo de atenção visual. Traduzido de Frintrop et al. (2010)

Neste projeto foi implementado uma adaptação do modelo de atenção visual proposto por Itti et al. (1998), sendo este um dos modelos mais conhecidos e tradicionais na literatura. Tal modelo tem como componente principal o mapa de saliência, o qual é constituído a partir da obtenção das características primitivas da imagem em diferentes escalas e aplicação de uma série de pirâmides Gaussianas.

O propósito do mapa de saliência é representar regiões salientes na cena e direcionar a seleção de regiões baseada na distribuição espacial da saliência (Pereira, 2007). É a partir deste que se obtém os marcadores que são utilizados no processo de segmentação descrito a seguir. Os marcadores correspondem às regiões que possuem valor de saliência igual ou superior ao valor de limiar saliência. Em outras palavras, representam as regiões que mais se destacam na imagem capturada.

Segmentação

A segmentação de imagens pode ser definida como uma técnica de processamento realizada com o objetivo de extrair informações de uma imagem por meio de métodos de classificação ou aglomeração, valendo-se da análise de características encontradas nos *pixels* daquela

como, por exemplo, níveis de cinza, textura, cor, entre outros (Khan and Ravi, 2013). A etapa de segmentação possui crucial relevância para o modelo, a medida que é por meio desta etapa que é realizada a identificação dos objetos de interação presentes no ambiente onde será realizada a interação.

O método de segmentação adotado deve garantir um funcionamento satisfatório na identificação de objetos de diferentes formatos e orientações, dadas as devidas proporções de tolerância a falhas e resultados indesejados.

O modelo adotou a técnica de segmentação *Watershed*. A técnica *Watershed*, ou inundação, é uma metodologia híbrida (Khan and Ravi, 2013), que mescla diversas características encontradas em outras técnicas de segmentação, sendo um método clássico da área de topografia e com uso na área de morfologia matemática (Klava, 2009; Audigier, 2004).

O funcionamento da abordagem *Watershed* é semelhante ao comportamento de uma inundação. Nesse cenário, uma quantidade regular de água inunda regiões por meio de furos que encontram-se nos mínimos regionais destas. A medida que tais regiões são inundadas, será chegado o momento onde duas bacias se encontrarão prestes a se fundir. Se é então formada uma barreira, impedindo a unificação das mesmas. Em determinado momento, apenas tais barreiras serão visualizadas no nível da água, determinando assim as linhas de *Watershed* (Klava, 2009; Audigier, 2004). É adotado o método de *Watershed* a partir de marcadores, onde são definidos os pontos específicos para o início das inundações. Os marcadores são escolhidos automaticamente por meio da utilização da técnica de atenção visual citado anteriormente.

Reconhecimento de Padrões

Reconhecer padrões é uma habilidade extremamente desenvolvida nos seres humanos. O ser humano é capaz de identificar, reconhecer e interpretar facilmente imagens, formas, sons ou conceitos. A facilidade que o ser humano possui para reconhecer padrões está diretamente vinculada aos estímulos pelos quais ele foi anteriormente exposto, bem como é capaz de evoluir através das suas experiências.

Entende-se por padrão as propriedades que possibilitam o agrupamento de objetos semelhantes dentro de uma determinada classe ou categoria, mediante a interpretação de dados de entrada, que permitam a extração das características relevantes desses objetos Tou and Gonzalez (1981) . Entende-se por classe de um padrão um conjunto de atributos comuns ao objetos de estudo.

As redes neurais artificiais (RNAs) já estão definitivamente consolidadas como uma alternativa à computação programada tradicional na solução de problemas (Ferro, 2013). Utilizou-se uma implementação de uma rede neural artificial denominada MLP (*Multi-Layer Perceptron*) ou rede *Perceptron* de múltiplas camadas, conforme Haykin (2001). Esta, por sua vez, consiste em arquitetura simples, que permite que a rede neural seja montada com uma camada de entrada, uma camada intermediária ou oculta e uma ca-

mada de saída.

Os segmentos obtidos são os sinais de entrada da rede MLP. Estes são classificados pela mesma a partir de um treinamento anteriormente realizado. Como saída da rede, tem-se o segmento classificado, sendo atribuído, assim, a sua respectiva ação.

Interação

A interação é realizada por meio do cálculo das diferenças de histogramas das imagens capturadas. A partir dos processos anteriormente explanados, é possível identificar as regiões de interesse que representam opções de ações e desta forma verificar se houve mudança de histograma nestas áreas. É a partir dessa mudança de histograma que se obtém o valor ΘIHC , valor este que pode variar de 0 a 1 e é utilizado como critério para a execução de ação. É observado se este ultrapassa um valor de limiar definido como mínimo para a computação de uma tentativa de interação válida.

O modelo de visão computacional descrito acima é utilizado durante este projeto de pesquisa como motor de processamento da nova aplicação sendo proposta. Ao longo deste projeto, o modelo será submetido a testes e adaptações a depender dos seus resultados quando inserido a uma nova aplicação que explora novos desafios e ambientes.

3.1.2 Câmera

Sabendo que a proposta do projeto consiste na interação residencial em ambientes claros e escuros, se fez necessária a modificação e adaptação de uma *webcam* comum para tornar possível a captura de imagens em ambientes com ou sem iluminação. Tal adaptação consiste na retirada do filtro infravermelho presente na câmera, além da utilização de LEDs infravermelhos Solanki and Desai (2011) e (Deshmukh et al., 2016). A Figura 3.7 exibe a estrutura desenvolvida durante o projeto para a *webcam*.



Figura 3.5: Estrutura construída para a *webcam*

A Figura 3.6 exibe imagens obtidas a partir da câmera adaptada. Nela observa-se um

comparativo entre a captura de imagem do ambiente iluminado com as luzes da residência e captura de imagem do ambiente com a ausência de iluminação, evidenciando o êxito na modificação realizada na *webcam*.



Figura 3.6: Comparativo entre a imagens capturadas utilizando a *webcam* modificada

3.1.3 Seleção de Ícones

Os ícones de interação foram selecionados e projetados utilizando como critério a sua função, simplicidade e significado intuitivo. A Figura 3.7 apresenta os cinco ícones selecionados para essa etapa do projeto, apresentando seus nomes e suas respectivas funções associadas. Tais ícones foram escolhidos a partir das características apresentadas na Seção 2.4, tendo em vista os itens mais presentes em residências. Vale ressaltar que o modelo é expansível e mais ícones podem ser facilmente incorporados à aplicação, adicionando a este mais funcionalidades.

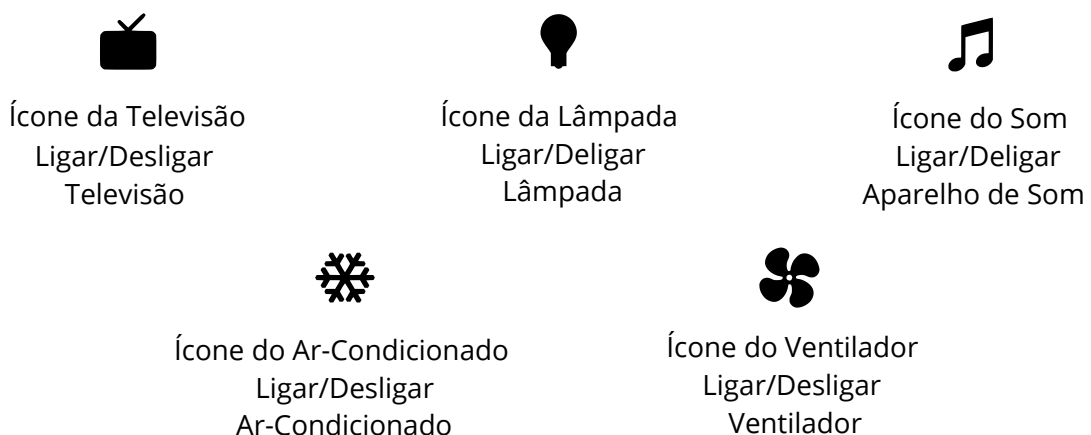


Figura 3.7: Ícones selecionados para a aplicação e suas respectivas funções

3.1.4 Plataforma de comunicação

A estrutura desenvolvida como solução para a comunicação e acionamento dos dispositivos residenciais buscou integrar o módulo Wi-Fi ESP8266 à plataforma Arduino. O circuito utilizado como base para a integração entre os módulos pode ser observado na Figura 3.8.

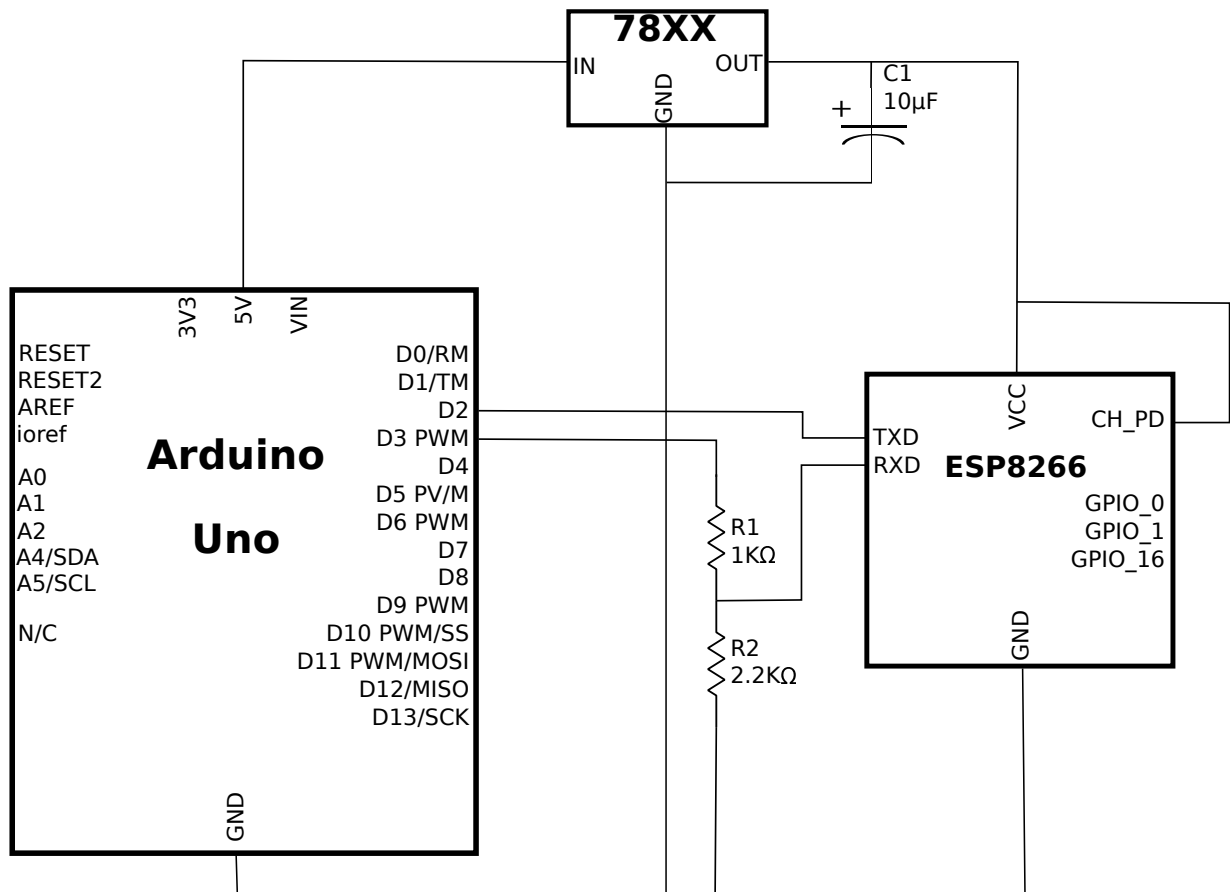


Figura 3.8: Circuito utilizado para a integração entre o ESP8266 e Arduino

Os componentes presentes no circuito consistem no Arduino Uno, o ESP8266, um regulador de tensão AMS1117 capaz de receber 5V e fornecer em seu terminal de saída uma voltagem de 3.3V, este foi utilizado aqui devido à saída de 3.3V presente no Arduino não oferecer potência suficiente para a alimentação eficiente do módulo ESP. Também está presente um capacitor para estabilizar o regulador de tensão e dois resistores funcionando como um divisor de tensão, visto que o o pino RX (Recepção serial) do ESP possui o nível de sinal de 3.3V e não pode ser ligado diretamente ao Arduino.

O protótipo desenvolvido a partir do circuito anteriormente apresentado e descrito está exibido na Figura 3.9. Este foi construído em uma *protoboard* e foi utilizado durante todos os testes realizados durante o processo de validação do método proposto. O protótipo em questão atende aos requisitos necessário para atingir os objetivos do projeto, e se diferencia do circuito original apenas pela adição de cinco leds representando as possíveis ações de interação. Estes foram incorporados ao modelo utilizando as saídas digitais de número quatro a oito do Arduino.

A seguir o código utilizado para configurar e inicializar o Arduino e o módulo ESP está exposto no Código 1. Este processo ocorre na função *setup*, a qual é invocada quando o dispositivo é ligado, sendo esta a primeira instrução a ser executada. No caso em questão, é durante a função *setup* que as saídas do Arduino são inicializadas, bem como o ESP8266 se conecta à rede Wi-Fi.

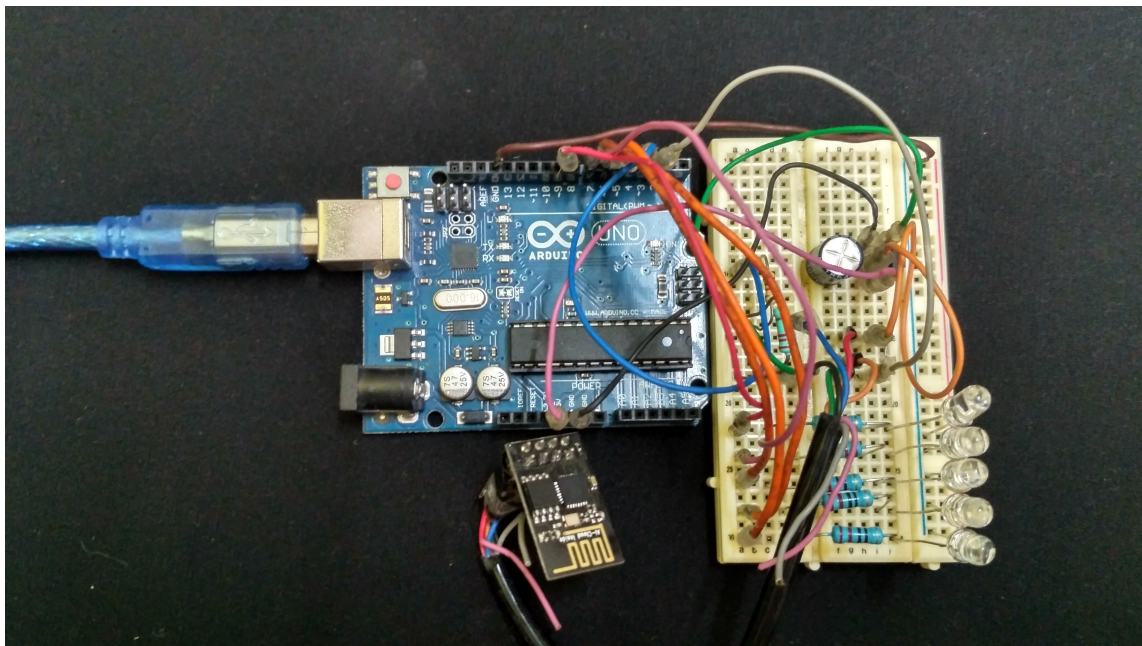


Figura 3.9: Protótipo construído utilizando o ESP8266 e Arduino

Algoritmo 1: Código de configuração de parâmetros do Arduino e ESP8266

```
1: void setup() {
2:   Serial.begin(9600);
3:   esp8266.begin(115200);
4:   for (int i = 4; i <= 8; i++) {
5:     pinMode(i, OUTPUT);
6:     digitalWrite(i, LOW);
7:   }
8:   sendData("AT+RST\r\n", 2000, DEBUG);
9:   sendData("AT+CWJAP=\"NomeDaRede\", \"Senha\"\r\n", 2000, DEBUG);
10:  sendData("AT+CWMODE=1\r\n", 1000, DEBUG);
11:  sendData("AT+CIPMUX=1\r\n", 1000, DEBUG);
12:  sendData("AT+CIPSTART=0, \"UDP\", \"0\", 0, 4445, 2\r\n", 1000, DEBUG);
13: }
14: }
```

As linhas 2 e 3 são responsáveis por determinar a velocidade de comunicação (*baud rate*) do Arduino, bem como do módulo ESP. O laço de repetição presente na linha 4 tem o intuito de determinar o comportamento dos pinos de número quatro a oito do Arduino como pinos de saída, responsáveis por acionar os leds. A função *sendData* é a instrução encarregada de enviar comandos AT para o ESP8266. O módulo é reiniciado na linha 9, na linha seguinte definem-se o nome da rede desejada para efetuar a conexão e sua respectiva senha. A instrução *AT+CWMODE* configura o modo de operação, no caso, o módulo está operando como um cliente Wi-Fi. A linha 12 habilita o ESP para aceitar múltiplas conexões, já a linha subsequente inicia uma conexão UDP na porta 4445, que será utilizada para a comunicação com o computador central.

O Código 2 exibe o código no método *loop* do Arduino. Esta é executada após a função

setup, e funciona como um laço de repetição, sendo executada até ser interrompida ou quando o dispositivo for desligado. Neste projeto, é na função *loop* que são tratadas as requisições enviadas pelo computador central através de uma conexão UDP através da porta 4445, além de executar a ação solicitada pelo mesmo.

Algoritmo 2: Código responsável por receber comandos e acionar saídas

```
1: void loop() {
2:     if (esp8266.available() && esp8266.find("+IPD,")) {
3:         delay(500);
4:         esp8266.find("acao=");
5:         int numeroAcao = (esp8266.read()-48);
6:         digitalWrite(numeroAcao, !digitalRead(numeroAcao));
7:         bool valorLeituraSaida = digitalRead(numeroAcao);
8:         String retornoServidor = (valorLeituraSaida == HIGH) ? "t":"f";
9:         String cipSend = "AT+CIPSEND=0,";
10:        cipSend += retornoServidor.length() + "\r\n";
11:        sendData(cipSend, 1000, DEBUG);
12:        sendData(retornoServidor, 1000, DEBUG);
13:        String closeCommand = "AT+CIPCLOSE=0\r\n";
14:        sendData(closeCommand, 1000, DEBUG);
15:        sendData("AT+CIPSTART=0,\"UDP\", \"0\", 0, 4445, 2\r\n", 1000, DEBUG);
16:    }
17: }
```

A linha 3 presente no código executa o comando *delay* do Arduino com o tempo de 500 milissegundos com o objetivo de aguardar a leitura de todos os dados recebidos pelo módulo ESP. A linha posterior realiza a busca pela mensagem “acao=” presente na mensagem enviada através conexão efetuada, já na instrução seguinte executa-se a leitura do número de ação enviado pelo computador central (a subtração por 48 é devido ao fato de que a função *read* retorna o valor decimal ASCII). A linha 6 através da função *digitalWrite* altera-se o valor da saída digital desejada, responsável pelo acionamento da mesma. Já nas linhas 7 e 8 realiza-se a leitura da situação atual da saída em questão, bem como é criada uma mensagem de retorno para o computador central, indicando se o led está ligado ou desligado. O comando para envio de mensagens para o computador central *AT+CIPSEND* é construído na linha 9 e 10. Por fim, encerra-se a conexão atual e é iniciada uma nova conexão UDP.

3.1.5 Reconhecimento de Funções Básicas para Interação Residencial

O método de Reconhecimento consiste em processar e classificar os objetos provindos da segmentação. Para que haja reconhecimento necessita-se de um treinamento prévio da

rede com imagens dos objetos de interação, para que a rede possa aprender e consequentemente identificar e classificar os objetos.

Para o treinamento da rede utilizou-se imagens representativas dos dispositivos acionados pela aplicação são eles: ventilador, ar condicionado, lâmpada, televisão e som. A figura 3.10 mostra as imagens de treinamento da rede neural.



Figura 3.10: Imagens utilizadas para o treinamento da rede neural

Para o treinamento é necessário um redimensionamento da imagem para 35x35 (padrão utilizado pela rede neural usada na aplicação) e binarização que equivale em transformar a imagem em apenas duas cores. A figura 3.10 demonstra as imagens já binarizadas em preto e branco. É necessário enfatizar que a rede foi treinada com dois objetos de ar condicionado como mostrado na figura 3.10, para que não houvesse favorecimento em nenhum método de segmentação, pois os dois métodos se comportam diferentes em relação ao ícone, no método *Watershed* a bacia inunda e preenche todo interior do objeto, já o método proposto de segmentação é mais delinear e o objeto toma uma forma mais parecida com o ícone utilizado na interação.

Depois do treinamento a rede está pronta para o reconhecimento dos objetos segmentados resultantes do plano de trabalho paralelo Particionamento da Cena em Regiões Relevantes, onde foi testado um novo método de segmentação e comparado com o método *Watershed*.

Foram realizados testes em ambientes diferenciados, com luminosidade, distância e ângulos diferentes da webcam. Vale ressaltar que o ambiente A-R é um recorte do ambiente A devido a limitações da distância e qualidade da imagem capturada. A figura 4.1 expõe os ambientes em que foram realizadas as experiências.

Em cada ambiente da figura 4.1 foram realizados testes com os objetos resultantes dos métodos de segmentação, os resultados serão mostrados na seção abaixo.

Resultados da etapa de Reconhecimento

A tabela 3.1 demonstra os resultados para cada ambiente testado e para cada método de segmentação, onde V.C representa o valor de classificação obtido para o segmento. Vale ressaltar que os objetos provindos da segmentação e da saliência tem papel crucial na classificação, se eles se distorcerem demasiadamente da imagem de treinamento a rede poderá não reconhecer o objeto ou reconhecer com um índice muito baixo.

Dada a tabela 3.1, pode-se concluir que o novo método de segmentação tem um grande potencial, pois na maioria dos casos resultou em melhores classificações, os valores em negrito da tabela mostram os resultados dos melhores valores de classificação, e o método



(a) Captura do ambiente A



(b) Captura do ambiente A-R



(c) Captura do ambiente B



(d) Captura do ambiente C



(e) Captura do ambiente D

Figura 3.11: Capturas dos ambientes

Tabela 3.1: Tabela de Classificação

Ambiente A				Ambiente A-R				Ambiente B				Ambiente C				Ambiente D			
Método proposto		Watershed		Método proposto		Watershed		Método proposto		Watershed		Método proposto		Watershed		Método proposto		Watershed	
Objetos	V.C	Objetos	V.C	Objetos	V.C	Objetos	V.C	Objetos	V.C	Objetos	V.C	Objetos	V.C	Objetos	V.C	Objetos	V.C	Objetos	V.C
	X		X		0,024		0,077		0,608		0,326		0,169		0,074		0,519		0,360
	0,422		X		0,184		0,153		0,939		0,929		0,846		0,848		0,486		0,602
	X		X		0,792		0,792		0,852		0,878		0,429		0,618		0,985		0,975
	X		X		0,999		0,998		0,997		0,998		0,951		0,114		0,999		0,999
	0,933		X		0,054		0,041		0,504		0,504		0,182		0,192		0,702		0,601

de segmentação proposto se saiu melhor em 12 vezes enquanto o *Watershed* teve valores mais altos em 7 casos.

É importante salientar que o método proposto conseguiu segmentar dois objetos no ambiente A em que a webcam estava em uma distancia longa do alvo e por isso a rede neural (RNA) conseguiu classificá-los corretamente mesmo com o valor baixo, já o *Watershed* não segmentou corretamente nenhum objeto do ambiente 1 como visto na tabela, O X representa os casos em que as imagens foram segmentadas de um forma muito distorcida e a Rede Neural Artificial (RNA) não conseguiu identificar o objeto.

Na tabela 3.1 no ambiente A-R recortado a rede errou a classificação no objeto da televisão proveniente do método de *Watershed*, o valor em vermelho salienta o erro da rede. Ao invés de reconhecer como televisão, a Rede Neural Artificial (RNA), reconheceu como ar condicionado, isso se dá pela imagem segmentada que se distorce muito da imagem de treinamento.

3.1.6 Reconhecimento de Tentativas de Interação

O último processo do modelo consiste na verificação de possíveis tentativas de interação com o sistema. A abordagem utilizada neste trabalho consiste em uma rotina continuamente em execução, responsável por capturar imagens e identificar a ocorrência de ação do usuário com algum dos ícones encontrados através dos processos anteriormente citados. A interação é reconhecida como uma mudança no valor de dissimilaridade entre os histogramas das regiões dos objetos identificados, aqui denominado por coeficiente IHC. O valor do coeficiente IHC é calculado através de uma adaptação da similaridade de cosseno, definida por:

$$dSim_{cos}(h_{ini}, h_{acq}) = 1 - \frac{\sum_{i=1}^n h_{ini_i} \cdot h_{acq_i}}{\sqrt{\sum_{i=1}^n (h_{ini_i})^2} \cdot \sqrt{\sum_{i=1}^n (h_{acq_i})^2}}, \quad (3.1)$$

onde $n = 255$, h_{ini} corresponde ao histograma do canal vermelho obtido a partir de uma captura e h_{acq} é o histograma do canal vermelho adquirido da imagem capturada posteriormente.

A rotina desenvolvida para a verificação de possíveis tentativas de interação pode ser observada no Código 3. Vale ressaltar a importância da linha 6, onde ocorre o cálculo da dissimilaridade dos histogramas das imagens por completo, este valor é utilizado com o objetivo de evitar um falso-positivo, ocasionado por variações de iluminação do ambiente.

Algoritmo 3: Código da rotina para verificação de ocorrência de interação

```
1: while (true) {
2:     int quantidadeOpcoes = 0;
3:     int acaoRealizada = 0;
4:     capturarImagem();
5:     histogramaAtualTotal = obterHistogramaImagemTotal();
6:     double dissimilaridadeTotal = dissimilaridade(
7:         histogramaAnteriorTotal, histogramaAtualTotal);
8:     for (int i = 0; i < quantidadeElementos; i++) {
9:         elementos[i].obterHistograma();
10:        double dissimilaridade = dissimilaridade(
11:            histogramaAnterior[i], elementos[i].getHistogramaVermelho());
12:        elementos[i].setCoeficienteIHC(dissimilaridade);
13:        if (elementos[i].getCoeficienteIHC() > 0.8
14:            && dissimilaridadeTotal < 0.2) {
15:            quantidadeOpcoes++;
16:            acaoRealizada = elementos[i].getClasse();
17:        }
18:        histogramaAnterior[i] = elementos[i].getHistogramaVermelho();
19:    }
20:    histogramaAnteriorTotal = obterHistogramaImagemTotal();
21:    if(quantidadeOpcoes == 1)
22:        enviarRequisicaoArduino(acaoRealizada);
23: }
```

O código consiste em laço de repetição continuamente executado, onde o primeiro passo é capturar uma imagem e, posteriormente, na linha 5, é obtido o histograma total da captura realizada. Em seguida, calcula-se a dissimilaridade entre o histograma da imagem atual e o histograma da última imagem adquirida anteriormente, esse valor é utilizado para a verificação de possíveis alterações externas no ambiente, como dito no parágrafo anterior. No laço presente na linha 7 calcula-se o histograma e dissimilaridade para cada elemento de interação identificado pelos processos anteriores presentes no modelo de visão computacional, bem como verifica-se a ocorrência de uma interação, executada na linha 11. Nas linhas 15 e 17 são armazenados respectivamente os histogramas de cada elemento e da imagem total, utilizados nos cálculos de dissimilaridades. A linha 18 impede que o sistema realize mais de uma ação simultaneamente. Já a linha 19 é responsável por enviar o comando executado para o Arduino.

Com o intuito de elucidar e auxiliar o entendimento do momento onde ocorre a identificação de interação, a Figura 3.12 apresenta dois exemplos de tentativas de interação. Nesta podem ser observados os histogramas obtidos, bem como os gráficos representando os coeficientes de interação obtidos calculados a partir da Equação 3.1 para a respectiva opção desejada.

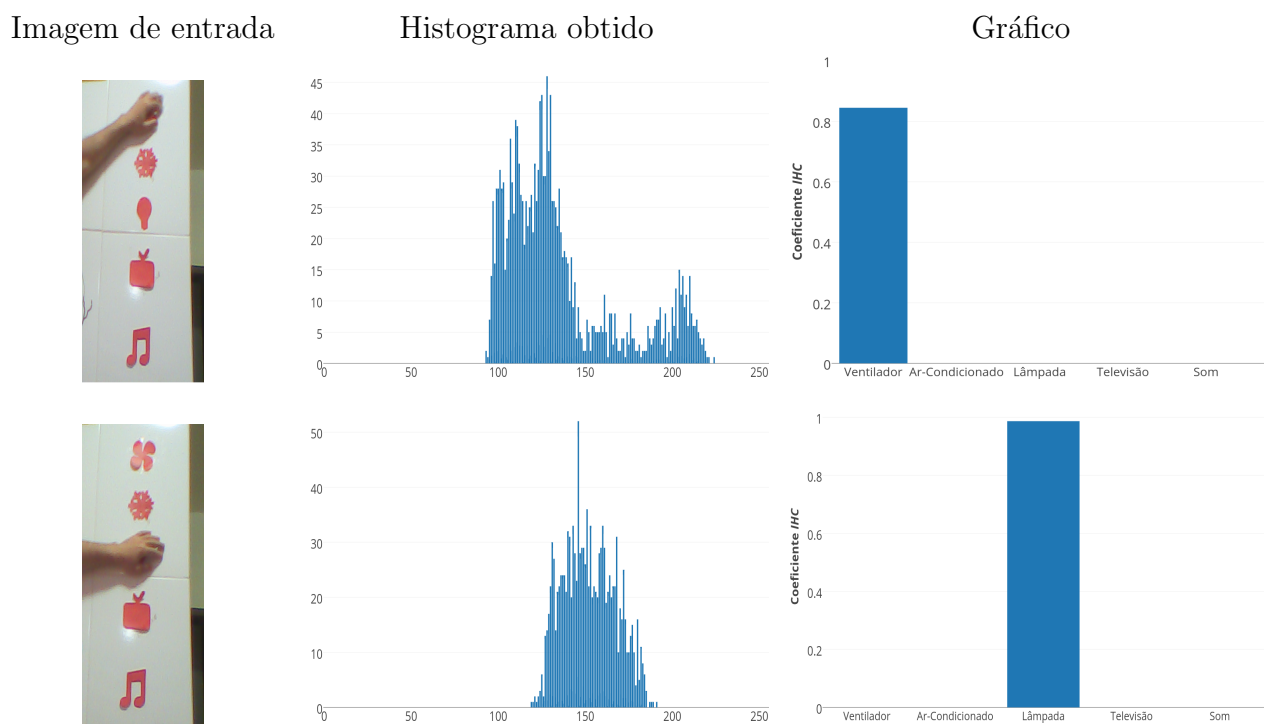


Figura 3.12: Tentativas de interação, histogramas obtidos e seus respectivos valores interação

O próximo Capítulo consiste na exposição e análise dos experimentos e testes realizados para a validação do funcionamento do modelo proposto.

Capítulo 4

Resultados

Buscando assegurar que o desempenho e robustez da aplicação atingiram níveis satisfatórios, testes e experimentos foram realizados. O processo de avaliação se deu através da tentativa de interações com o sistema e a verificação das respostas obtidas como resultados.

Os testes foram executados em diferentes ambientes residenciais, com variações de luminosidade, fonte de luz e direção de incidência destas sob os ícones de interação. A variação de luminosidade é um fator de grande importância, pois se mostrou determinante nos resultados obtidos, interferindo no nível de sensibilidade e tentativas de interação. Foram definidos valores para os parâmetros do sistema, que estendem-se por todas as etapas do modelo, isto é, desde a fase de pré-processamento até a etapa de interação, estes permaneceram inalterados nos diferentes ambientes. Assim, buscou-se aferir a robustez do modelo proposto, de maneira a verificar se o seu funcionamento se mostrou correto sem a necessidade de configurações ou calibrações. Os ambientes utilizados para os testes podem ser observados na Figura 4.1.

No total, foram realizadas 530 interações em ambientes iluminados, sendo divididas em 106 tentativas de interação para cada opção disponível. Para a análise dos resultados obtidos foi utilizada uma matriz de confusão, apresentada na Tabela 4.1. Todos os testes foram agrupados e dispostos na mesma e o conjunto de classes foi dividido de acordo com as opções de interação presentes nos testes, com adição da coluna representativa de interações mal-sucedidas (IMS) e acurácia individual. É importante salientar que as IMSs não são consideradas como uma classe, mas sim a ausência de execução da respectiva ação por parte da aplicação. Adicionalmente, a acurácia geral foi calculada, obtendo um resultado de 96.04%.

Tabela 4.1: Matriz de confusão e acurácias individuais obtidas em ambientes claros

Ações	Ventilador	Lâmpada	Ar-Condicionado	TV	Som	IMS	Acurácia
Ventilador	103	0	0	0	0	3	99.41%
Lâmpada	0	105	0	0	0	1	99.80%
Ar-Condicionado	0	0	101	0	0	5	99.03%
TV	0	0	0	103	0	3	99.41%
Som	0	0	0	0	97	9	98.26%
IMS	0	0	0	0	0	0	0%



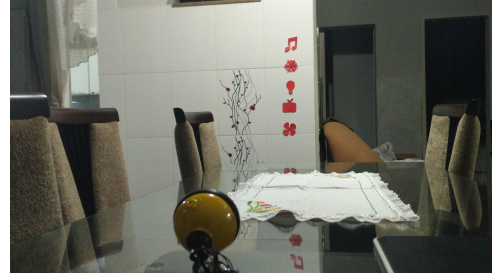
(a) Ambiente A



(b) Ambiente B



(c) Ambiente C



(d) Ambiente D

Figura 4.1: Ambientes utilizados para os testes

A Figura 4.2 exibe um gráfico referente as respostas de tentativas de interação realizadas durante um dos testes em ambientes iluminados. No mesmo estão representadas as cinco opções de interação presentes no sistema. A cada tentativa de interação com uma das opções, é disparado uma linha no gráfico referente a mesma. O limiar mínimo de reconhecimento de uma interação (função de ativação) ficou definido como θ_{IHC} de 0.8, isto é, apenas é reconhecida como uma tentativa de interação válida as linhas que atingem valores superiores a tal limiar.

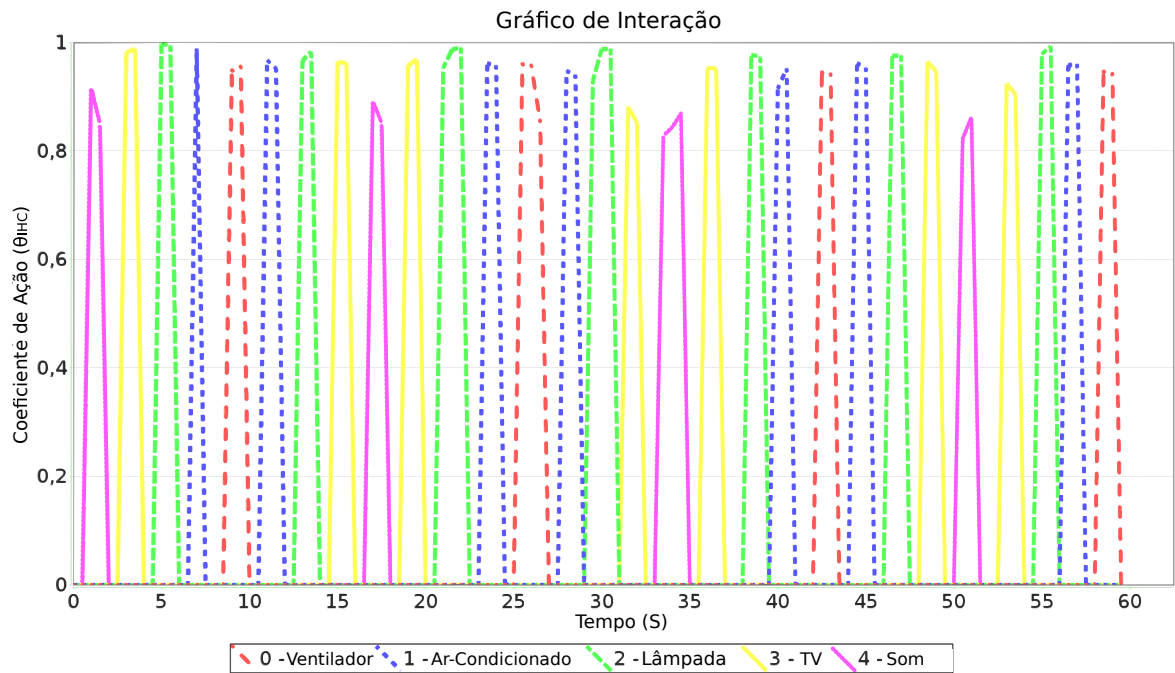


Figura 4.2: Gráfico de interação em um ambiente iluminado e os coeficientes IHC obtidos

De acordo com os objetivos deste trabalho, que se referem ao funcionamento e operabilidade do modelo em ambientes em condições noturnas, foram realizados testes específicos para tal configuração. As condições dos ambientes utilizados podem ser visualizadas na Figura 4.3, onde a webcam modificada capturou as imagens com auxílio dos leds infravermelhos.

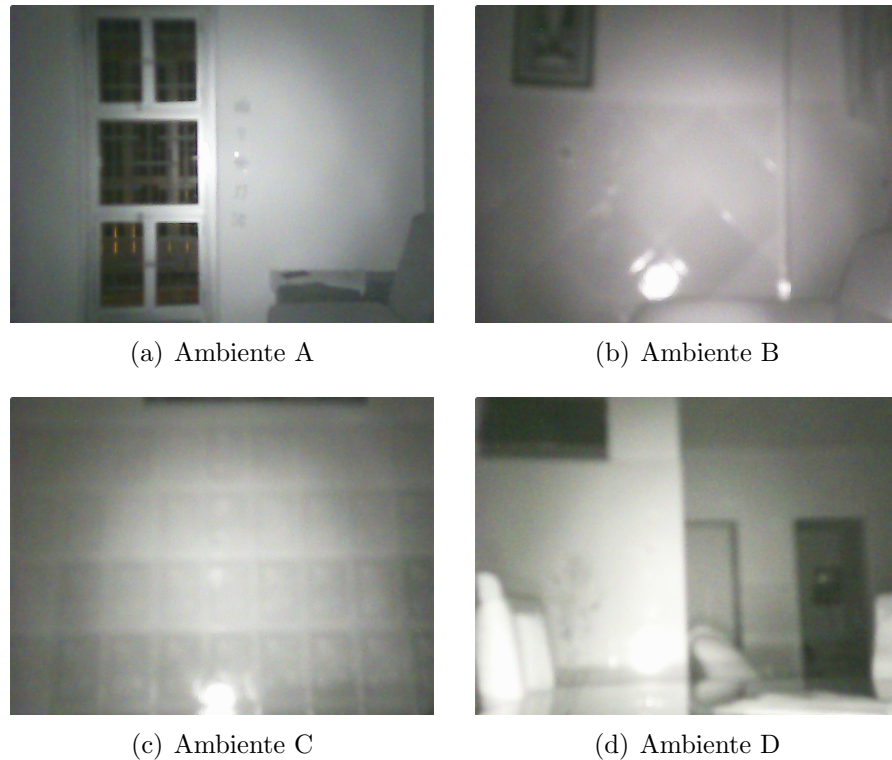


Figura 4.3: Imagens capturadas pela webcam utilizando luz infravermelha

Nas condições acima citadas, foram executadas o total de 205 interações, distribuídas em 41 tentativas por ícone. Como realizado anteriormente, os resultados foram computados em uma matriz de confusão, com a adição das interações mal-sucedidas (IMS) e acurácias individuais. Este conjunto de testes obteve uma acurácia geral de 96.59%. A decisão de dispor os testes em matrizes separadas partiu da necessidade de destacar o funcionamento do modelo em ambientes escuros, um ponto de grande importância e relevância do projeto.

Tabela 4.2: Matriz de confusão e acurácias individuais obtidas em ambientes escuros

Ações	Ventilador	Lâmpada	Ar-Condicionado	TV	Som	IMS	Acurácia
Ventilador	38	0	0	0	0	3	98.50%
Lâmpada	0	41	0	0	0	0	100.0%
Ar-Condicionado	0	0	40	0	0	1	99.49%
TV	0	0	0	39	0	2	99.00%
Som	0	0	0	0	40	1	99.50%
IMS	0	0	0	0	0	0	0%

A seguir, na Figura 4.4, é exposto um gráfico de interações retirado de um dos testes executados em ambientes escuros. Vale ressaltar que para os testes em ambiente sem iluminação nenhum parâmetro ou variável do método de verificação de ocorrência de

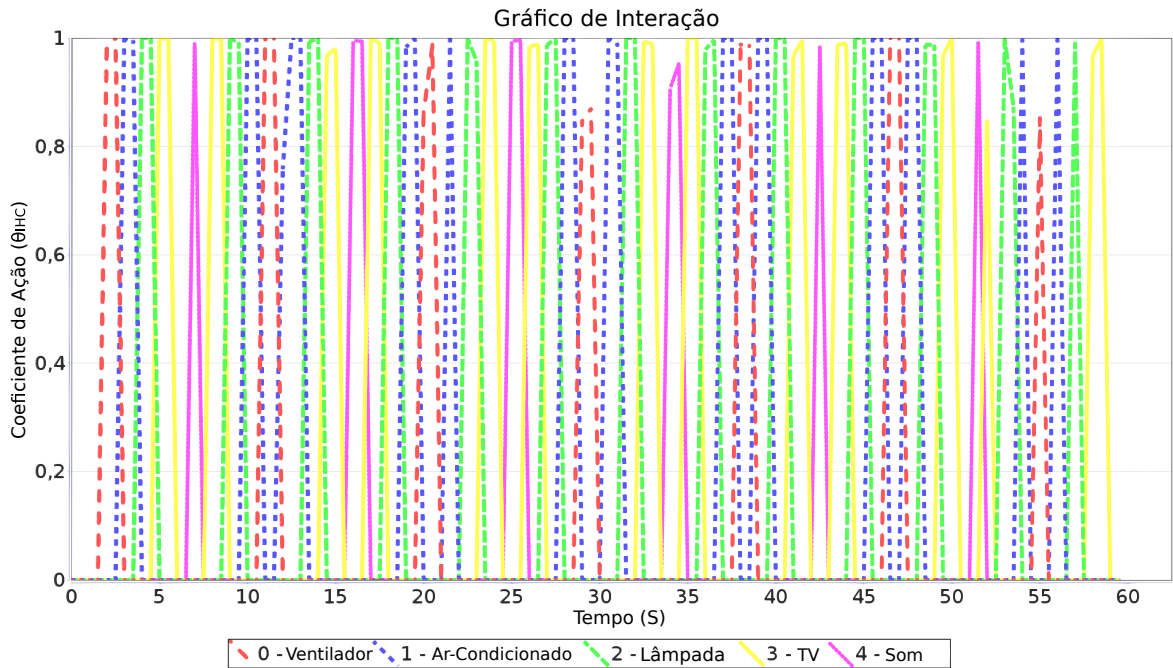


Figura 4.4: Gráfico de interação em um ambiente sem iluminação e os coeficientes IHC obtidos

interação foi alterado. O valor do limiar de reconhecimento de uma interação, ou θ_{IHC} , permaneceu 0.8.

Os experimentos apresentaram resultados consistentes nas várias condições e configurações de luz e cenários, atingindo os valores satisfatórios. Um fator importante é que nenhuma ação errada foi disparada, contribuindo com a eficiência do modelo utilizado. Além disso, como dito anteriormente, as interações mal-sucedidas são definidas como uma tentativa de interação do usuário que não foi reconhecida pelo sistema, e não uma ação disparada arbitrariamente.

Por fim, como pode ser constatado nas Figuras 4.2 e 4.4, as ações de interação foram detectadas com margens confortáveis, obtendo valores do coeficiente θ_{IHC} próximo de 1 para as ações realmente desejadas e zero para as demais opções, demonstrando a robustez do modelo proposto.

Capítulo 5

Conclusões

Neste trabalho, foi proposto um modelo de interação baseado em técnicas de inteligência artificial e visão computacional para ambientes residenciais sem uso de sensores ou controladores especiais. O modelo proposto foi capaz de detectar e reconhecer os ícones de interação corretamente em ambientes diversos, com escalas variadas entre a câmera e os ícones de destino, e diferentes condições de luminosidade no ambiente, dispensando alterações de parâmetros ou qualquer tipo de calibração.

Com base nos testes realizados, resultados obtidos por meio da análise da matriz de confusão e dos gráficos de interações, foi observado que o modelo proposto encontra-se funcional e alcança o seu objetivo, reforçando a robustez e confiança no modelo proposto.

Além disso, o sistema mantém a operabilidade em condições diurnas e noturnas, utilizando apenas uma câmera comum, dispensando sensores ou outros equipamentos especiais, fornecendo uma solução eficiente de baixo custo, fácil aprendizagem e uma forma mais natural de interação apenas utilizando gestos. Essas características foram possíveis pelo uso de Inteligência Artificial e técnicas de Visão Computacional, direcionando a complexidade e as exigências cognitivas ao modelo e não aos usuários.

É válido de nota a proposição neste trabalho de um método de segmentação próprio, tendo o seu desenvolvimento iniciado e obtendo resultados preliminares eficientes, com menor custo computacional que o *Watershed* utilizado até então, alcançando taxas de reconhecimento equivalentes. É de interesse a continuidade do desenvolvimento deste método, incorporando melhorias e aperfeiçoando-o.

Adicionalmente, como já explicitado neste e nos demais planos componentes deste projeto, o atual modelo pode ser incorporado a outras aplicações, podendo assim oferecer todas as vantagens de interatividade a baixo custo para, por exemplo, aplicações educacionais, exposições em museus com interatividade, entre outras.

Em conclusão, o modelo demonstrou resultados satisfatórios e relevantes, bem como foi exitoso em sua proposta de um modelo alternativo que possibilita a interação por meio de gestos, dispensando dispositivos especiais ou sensores, oferecendo uma nova opção para controlar uma residência. Sendo este, como já citado na introdução, publicado na 19ª Conferência Internacional sobre Interação Homem-Computador (HCI Internacional 2017)

realizada em Vancouver, Canadá, no período de 9 a 14 de julho de 2017, com o título de Human-House Interaction Model Based on Artificial Intelligence for Residential Functions (Santos et al., 2017). Tal publicação consta como um capítulo do livro HCI International 2017 - Posters' Extended Abstracts.

Como trabalhos futuros, melhorias no método para detectar interações podem ser investigadas, aplicando abordagens mais inteligentes. Deverá ser implementado também o ajuste automático de parâmetros a partir do reconhecimento do ambiente. Adicionalmente, mais opções de ícones deverão ser incorporadas ao modelo. Vale lembrar que o modelo de visão aqui proposto e utilizado é expansivo e pode ser aplicado, mediante adaptações, a outras situações diferentes de um ambiente residencial.

Referências Bibliográficas

- Aggarwal, J. K. and Q. Cai (1997). Human motion analysis: A review. In *Nonrigid and Articulated Motion Workshop, 1997. Proceedings., IEEE*, pp. 90–102. IEEE.
- Ahmed, A., I. Siddiqui, and M. M. Khan (2016). Ip network based home automation system using avr microcontroller. In *Computing, Electronic and Electrical Engineering (ICE Cube), 2016 International Conference on*, pp. 14–21. IEEE.
- Ahmim, A., T. Le, E. Ososanya, and S. Haghani (2016). Design and implementation of a home automation system for smart grid applications. In *2016 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 538–539. IEEE.
- Apoorva Bharambe, Divya Chanekar, D. N. and A. B. Vitekar (2015). Automatic hand gesture based remote control for home appliances. *International Journal of Advanced Research in Computer Science and Software Engineering*.
- Audigier, R. M. M. (2004). *Segmentação e Visualização Tridimensional Interativa de Imagens de Ressonância Magnética*. Ph. D. thesis, UNIVERSIDADE ESTADUAL DE CAMPINAS.
- Baig, I., C. Muzamil, and S. Dalvi (2016). Home automation using arduino wifi module esp8266.
- Bartos, F. (2013). Java Webcam Capture. Disponível em: <https://github.com/sarxos/webcam-capture>.
- Benicasa, A. X. (2013). *Sistemas computacionais para atenção visual Top-Down e Bottom-up usando redes neurais artificiais*. Ph. D. thesis, Universidade de São Paulo.
- Blake, J. (2012). *Natural User Interfaces in . Net*. Manning Pubs Co Series. Manning Publications Company.
- Celebre, A. M. D., A. Z. D. Dubouzet, I. B. A. Medina, A. N. M. Surposa, and R. C. Gustilo (2015). Home automation using raspberry pi through siri enabled mobile devices. In *Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), 2015 International Conference on*, pp. 1–6. IEEE.

- Chan, M., C. Hariton, P. Ringear, and E. Campo (1995). Smart house automation system for the elderly and the disabled. In *Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century., IEEE International Conference on*, Volume 2, pp. 1586–1589. IEEE.
- Correia, A. C. d. C. (2014). Interação gestual em ambiente residencial: estudo, concepção e desenvolvimento.
- Creston (2016). Creston product overview. 2016.
- Delazeri, H. W., S. de Castro Bertagnolli, P. N. Hubler, and F. d. O. de Freitas (2017). Automação de ambientes utilizando a plataforma arduino e conexão wifi. *Anais do Salão de Iniciação Científica e Tecnológica e Salão de Extensão do IFRS-Câmpus Canoas 1* (6).
- Deshmukh, M. R., A. Bange, A. Nerkar, and S. Mane (2016). Automatic hand gesture based remote control for home appliances.
- Durfee, W. (2011). Arduino microcontroller guide. minnesota: University of minnesota.
- e Woods R. E. Gonzalez, R. C. (2010). *Processamento Digital de Imagens - 3a ed.* São Paulo: Pearson.
- Erol, A., G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly (2005). A review on vision-based full dof hand motion estimation. In *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, pp. 75–75. IEEE.
- et al. Dias, D. R. C. (2013). Desenvolvimento de Aplicações com Interface Natural de Usuário e Dispositivos PrimeSense como Meio de Interação para Ambientes Virtuais.
- Ferro, L. (2013). Aplicação da rede neural mlp (multilayer perceptron) em indústria de pisos e revestimentos do pólo cerâmico de santa gertrudes-sp.
- Frintrop, S., E. Rome, and H. I. Christensen (2010). Computational visual attention systems and their cognitive foundations: A survey. *ACM Transactions on Applied Perception (TAP)* 7(1), 6.
- Grando, N. (2005). Segmentação de Imagens Tomográficas Visando a Construção de Modelos Médicos.
- Haykin, S. S. (2001). *Redes neurais*. Bookman.
- Itti, L., C. Koch, and E. Niebur (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence* (11), 1254–1259.

- Jain, A. K. and C. Dorai (1997). Practicing vision: Integration, evaluation and applications. *Pattern Recognition* 30(2), 183–196.
- Java, . (2014). Java. Disponível em: [http :
//www.java.com/pt_BR/download/faq/whatis_java.xml/..](http://www.java.com/pt_BR/download/faq/whatis_java.xml/..)
- Javafree, . (2014). Javafree. Disponível em: [http :
//javafree.uol.com.br/artigo/871497/Tutorial – Java – 3 – Orientacao – a –
Objetos.html#ixzz3DTfalq4F..](http://javafree.uol.com.br/artigo/871497/Tutorial%20-%20Java%20-%203%20-%20Orientacao%20a%20Objetos.html#ixzz3DTfalq4F..)
- Khan, A. and S. Ravi (2013). Image segmentation methods: A comparative study.
- Klava, B. (2009). *Segmentação interativa de imagens via transformação watershed*. Ph. D. thesis, Master’s thesis, Instituto de Matemática e Estatística-Universidade de São Paulo.
- Kumar, S. and S. Solanki (2016). Voice and touch control home automation. In *Recent Advances in Information Technology (RAIT), 2016 3rd International Conference on*, pp. 495–498. IEEE.
- Lee, K. Y. and J. W. Choi (2003). Remote-controlled home automation system via bluetooth home network. In *SICE 2003 Annual Conference*, Volume 3, pp. 2824–2829. IEEE.
- Lutron (2016). Radiora 2 save energy with wireless total home control from lutron. 2016.
- Mehta, M. (2015). Esp 8266: a breakthrough in wireless sensor networks and internet of things. *International Journal of Electronics and Communication Engineering & Technology (IJECEET)* 6(8), 07–11.
- Mittal, Y., P. Toshniwal, S. Sharma, D. Singhal, R. Gupta, and V. Mittal (2015). A voice-controlled multi-functional smart home automation system. In *2015 Annual IEEE India Conference (INDICON)*, pp. 1–6. IEEE.
- Morales, D., T. M. Centeno, and R. C. Morales (2003). Extração Automática de Marcadores Anatômicos no Desenvolvimento de um Sistema de Auxílio ao Diagnóstico Postural por Imagens.
- Morgan, J. (2008). Técnicas de Segmentação de Imagens na geração de Programas para Máquinas de Comando Numérico.
- Morshed, N. M., G. Muid-Ur-Rahman, M. R. Karim, and H. U. Zaman (2015). Micro-controller based home automation system using bluetooth, gsm, wi-fi and dtmf. In *Advances in Electrical Engineering (ICAEE), 2015 International Conference on*, pp. 101–104. IEEE.

- Muratori, J. R. and P. H. Dal Bó (2011). Capítulo I automação residencial: histórico, definições e conceitos. *O setor elétrico*, 70.
- Pannase, D. and S. Ugale (2014). To analyze hand gesture recognition for wirelessly electronic device control. *IOSR Journal of Electronics and Communication Engineering Vol. 09*.
- Pereira, E. T. (2007). Atenção visual bottom-up guiada por otimização via algoritmos genéticos.
- Prabhuraj, R. and B. Saravanakumar (2014). Gesture controlled home automation for differently challenged people. *International Journal of Research in Electronics 2*, 1–6.
- Raut, A., V. Singh, V. Rajput, and R. Mahale (2012). Hand sign interpreter. *HAND 1* (2), 19–25.
- Ricquebourg, V., D. Menga, D. Durand, B. Marhic, L. Delahoche, and C. Loge (2006). The smart home concept: our immediate future. In *2006 1st IEEE international conference on e-learning in industrial electronics*, pp. 23–28. IEEE.
- Santos, B. F. L., I. B. de Andrade Santos, M. J. M. Guimarães, and A. X. Benicasa (2017). Human-house interaction model based on artificial intelligence for residential functions. In *International Conference on Human-Computer Interaction*, pp. 349–356. Springer.
- Santos, I. B. A., B. F. L. Santos, L. A. Fonseca Sobrinho, and A. X. Benicasa (2016). Interação natural dinâmica com imagens projetadas. *XIII Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*, 733–742.
- Silva, B. R. S. d. (2014). Sistema de automação residencial de baixo custo para redes sem fio.
- Solanki, U. V. and N. H. Desai (2011). Hand gesture based remote control for home appliances: Handmote. In *Information and Communication Technologies (WICT), 2011 World Congress on*, pp. 419–423. IEEE.
- Sombra, L. G. (2016). Automação residencial para controle de iluminação, segurança e monitoramento de temperatura usando o arduino mega.
- Sumanta Dey, Amit Das, A. M. and D. Dutta (2015, feb). Gesture controlled home automation. *International Journal of Emerging Engineering Research and Technology*.
- TechinBrazil (2015). Integradores da automação residencial. 2015.
- Tou, J. T. and R. C. Gonzalez (1981). Pattern recognition principles.

- Treisman, A. M. and G. Gelade (1980). A feature-integration theory of attention. *Cognitive psychology* 12(1), 97–136.
- Valli, A. (2007). Natural interaction. *Recuperado de <http://www.naturalinteraction.org/images/whitepaper.pdf>*.
- Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic bulletin & review* 1(2), 202–238.

ANEXO I

Trabalho de Santos et al. (2017), submetido e aprovado para a 19^a Conferência Internacional sobre Interação Homem-Computador (HCI Internacional 2017) realizada em Vancouver, Canadá, no período de 9 a 14 de julho de 2017. Tal publicação consta como um capítulo do livro HCI International 2017 - Posters' Extended Abstracts.

Human-House Interaction Model Based on Artificial Intelligence for Residential Functions

Brendel Francisco Lima Santos, Iury Batista de Andrade Santos,
Maurício José Miranda Guimarães, and Alcides Xavier Benicasa

Department of Information Systems - DSI,
Federal University of Sergipe - UFS,
Itabaiana-SE, Brazil
{brendelsantos,iurybas,mauriciojosemirandaguimaraes}@gmail.com
alcides@ufs.br

Abstract. The scenarios as smart homes and its devices requires novel ways to perform interactive actions. In this work we explore and develop a model to interact, in a natural, easy learning and intuitive manner, with a smart home, without use special sensors or another controllers, based on interpretation of complex context images captured with a trivial camera. We use artificial intelligence and computer vision techniques to recognize action icons in a uncontrolled environment and identify user interact actions gestures. Our model connects with well know computational platforms, which communicate with devices and another residential functions. Preliminary tests demonstrated that our model fits well for the objectives, working in different conditions of light, distance and ambiances.

Keywords: Human-House interaction · Artificial intelligence · Computer vision · Residential automation.

1 Introduction

Forms of interaction between humans and computers have been evolved to new and diverse objects and environments of our day-to-day life. Advances in technologies, as miniaturization of devices, tools for no wired communications, and others, have provided the design of applications for smart homes, turning casual activities more simple, comfortable and intuitive. Residential automation is composed by a set of sensors, equipments, services and diverse technological systems integrated with the aim of assist basic necessities of security, communication, energetic management and housing comfort [1–3].

Those scenarios demand ways of interaction more dynamic, natural and easy learning. Several works presents novel interaction solutions based on voice commands, facial expressions, as well capture and interpretation of gestures, the last a point of investigation of this work. Systems that use free gestures interpret actions naturally realized by persons to communicate, giving the possibility of a more easy and intuitive way of interact to users, reducing the cognitive

overload of information, training and learning [4]. Such type of interaction was perceived by researches in Human-Computer Interactions as relevant since the first investigations [5].

Therefore, we propose a model of interaction for smart homes without use of sensors or controllers, based on artificial intelligence techniques and computer vision. These methods are applied to recognizing actions of interaction with targets dynamically distributed in a residence, using capture of images from a trivial camera, in a manner similar to a touch area. This approach aims a more cheap, efficient and simple manner of interact with a smart home and its electric devices, using representative symbols and artificial intelligence techniques, recognizing interactions from the user. Our work connect these techniques with well known computational platforms, which communicate with devices and another residential functions.

This work is organized as follow: in the next section, we present related works in residential automation. Our model, the techniques used and methodology is explicated in the section Methodology. Experiments and results are showed in section Experiments and Results. Finally, we conclude discussing the results and possible future improvements and researches to be explored.

2 Related Work

Capture of the gestures can be performed through the use of special devices, such as sensors, accelerometers and gloves. These accessories facilitate interpretation, however, the total cost of the system is increased [6]. In this sense, the work developed in Bharambe [7] presents a system composed of an accelerometer to capture the gestures, a microcontroller responsible for identifying the information collected, a transmitter and an infrared receiver. It is worth mentioning that there is a limitation regarding distance, in addition to the fact that the infrared transmitter and receiver must be fully aligned.

In contrast to the use of special accessories, the methods based on computer vision require only a computer camera to capture images and automatically interpret complex scenes [8]. Therefore, Sumanta [9] proposed a work where the capture and interpretation of gestures is performed using the camera of a smartphone. A binary signal is generated from the captured image, which is graphically analyzed and transmitted using bluetooth to a control board responsible for activate the devices in the residence.

Still referring to methods based on computer vision, Pannase [10] developed a research that focused on quickly detecting hand gestures. For this, algorithms of segmentation and detection of the positioning of the fingers were used, and, finally, the classification of the gesture through the neural network. Another aspect of this study is the limited amount of gestures, as well as the hand should be positioned exactly in front of the camera.

In this work, we propose a model of interaction between a human and a residential environment using gestures, dispensing especial sensors or another devices, providing a interaction solution more natural and intuitive to control

the functions of a smart home as, for example, switch on and switch off electronic devices, lamps, open and close doors and windows etc. The system must be able to operate in both daytime and nighttime conditions. For this, a webcam was adapted to see infrared light.

In the next section will be presented the main characteristics of the proposed model.

3 Metodology

For the development of this project, techniques of artificial intelligence and computer vision, as well complex contexts images, have been adopted to identify interaction targets and actions of the user with the ambiance.

Our model is composed as presented in Figure 1. An application located in a main computer is responsible for the execution of all procedures involving the techniques of artificial intelligence and computer vision. These procedures are applied upon a image captured from a webcam located in the environment, directed to where the target icons are placed. The action icons are disposed in the ambiance as desired by the user. The application in the main computer identifies if a user interaction action occurred. If positive, the respective action command is send to the control board, which is connected with the devices and another interactive functions of the residence.

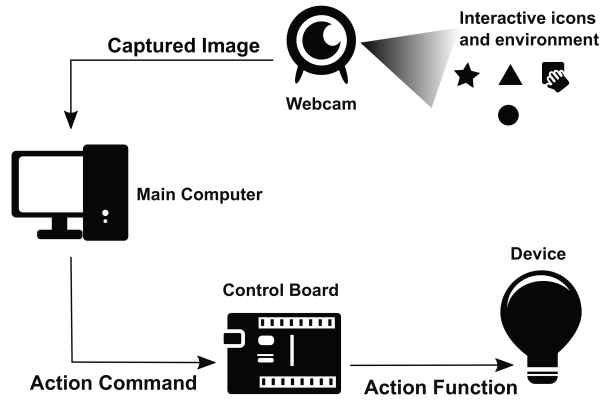


Fig. 1. Scheme of the model being proposed in this work, with it components and flux of information

To illustrate the operation of the proposed model was used a miniature of a residence. Is possible notice the triggering of a light as the interaction with an object occurs, as seen in the Figure 2.



Fig. 2. Illustration of model working in a miniature

3.1 The Computer Vision Model

The application in the main computer incorporate a model developed in previous researches made by our research team. This model integrate different techniques of artificial intelligence, computer vision and image processing to discovery and classify specific objects (in this case, target icons of interaction) in a complex and low controlled environment. From a captured image of the ambiance, a pre-processing is performed for saturation of red-colored pixels, then we apply visual attention to identify interest areas in the image and discard any other visual information. In this step we have modified the classic method of visual attention as proposed in Itti [11], simplifying it to respond only for color stimulus. The areas of interest are marked as seeds, which are used as input in the segmentation process. For segmentation we have used the method Watershed, as developed by Klava [12]. The segmentation step is responsible for generating individual elements of the target icons from the input image. These elements are classified using neural networks, were a action of interaction is attributed in respect with the type of each icon.

In follow, the interaction routines are started, continuously capturing images and searching for possible interact actions performed by the user. Interactions are recognized as changes in the coefficient of dissimilarity between the histograms of the regions where icons of interaction has been found. The dissimilarity $dSim_{cos}$ is calculated using an adaptation of the cosine similarity defined by:

$$dSim_{cos}(h_{ini}, h_{acq}) = 1 - \frac{\sum_{i=1}^n h_{ini_i} \cdot h_{acq_i}}{\sqrt{\sum_{i=1}^n (h_{ini_i})^2} \cdot \sqrt{\sum_{i=1}^n (h_{acq_i})^2}}, \quad (1)$$

where $n = 255$, h_{ini} corresponds to the histogram of the red channel obtained from an initial capture and h_{acq} is the histogram of the red channel acquired from subsequent captured images.

The scheme of the computer vision model is show in Figure 3.

3.2 The Human-House Interaction Model

The Human-house interaction model is composed as previously show in Figure 1. The interactive icons placed in the environment was designed considering its function, simplicity and intuitive meaning. The icons are tested in diverse ambiances, with different conditions of light, distance from the webcam, and elements in the surrounds, to assure its correct recognizability by the computer

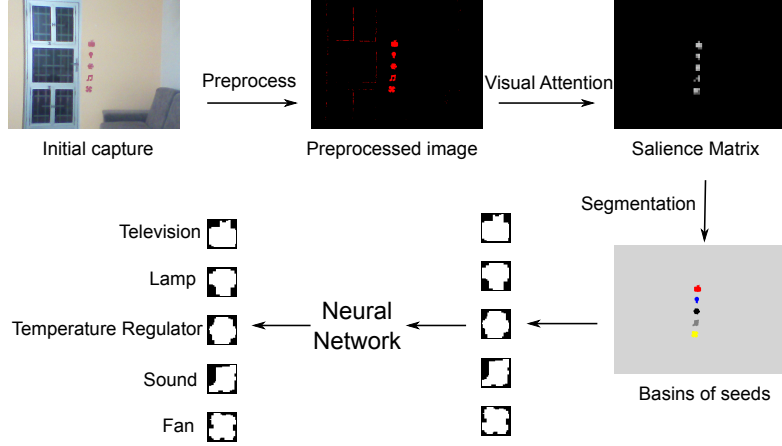


Fig. 3. Scheme of the computer vision model implementing artificial intelligence techniques

vision model. The interactive functions are elected by its pertinence and utility. In Figure 4 there is five icons of interaction used by our model, presented with their names and related functions.

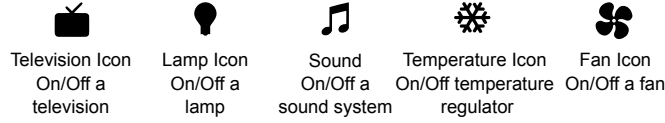


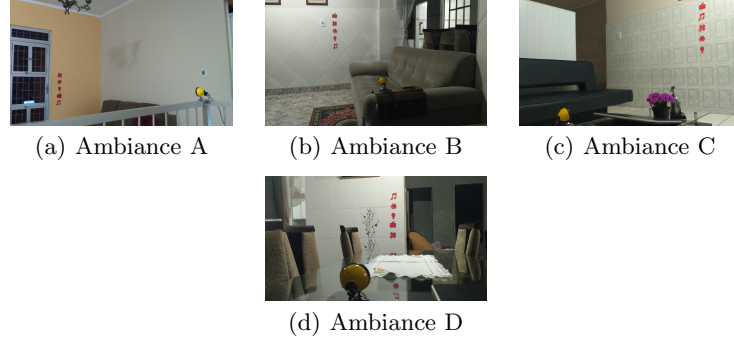
Fig. 4. Icons set of actions of interaction

Its is pertinent mention that our model is expansible and more icons can be introduced in the application.

4 Experiments and Results

The experiments were realized in several residential environments varying the positioning of the camera and with different levels of luminosity, including total absence of light. The process occurred through the attempt of interactions and the answers obtained as results. The environments where the tests were performed can be seen as in Figure 5.

In total, 530 interactions were executed in ambiances with luminosity, being divided into 106 interactions attempts per option. The results in these conditions are presented through a confusion matrix, unsuccessful interactions (UI) and individual accuracy, as seen in Table 1. In addition, the total accuracy was calculated, with a result of 96.04%.

**Fig. 5.** Environments used in the tests**Table 1.** Confusion matrix and individual accuracy of interaction in environments with luminosity

Actions	Fan	Lamp	Temperature	TV	Sound	UI	Accuracy
Fan	103	0	0	0	0	3	99.41%
Lamp	0	105	0	0	0	1	99.80%
Temperature	0	0	101	0	0	5	99.03%
TV	0	0	0	103	0	3	99.41%
Sound	0	0	0	0	97	9	98.26%
UI	0	0	0	0	0	0	0%

According with the objectives of this work, which refers to operation in nocturnal conditions, have been realized specific tests for these settings. Were executed the total of 205 interactions, distributed in 41 attempts per option. As done previously, the results are showed through confusion matrix, unsuccessful interactions (UI) and individual accuracy. The Table 2 exposes the results. In this case, the total accuracy obtained was 96.59%.

Table 2. Confusion matrix and individual accuracy of interaction in dark environments

Actions	Fan	Lamp	Temperature	TV	Sound	UI	Accuracy
Fan	38	0	0	0	0	3	98.50%
Lamp	0	41	0	0	0	0	100.0%
Temperature	0	0	40	0	0	1	99.49%
TV	0	0	0	39	0	2	99.00%
Sound	0	0	0	0	40	1	99.50%
UI	0	0	0	0	0	0	0%

In Figure 6 is presented a graphic where interaction actions can be seen with its respective coefficient of dissimilarity, obtained using the Equation 1, representing the human-house interaction value (θ_{HHI}). Only θ_{HHI} values above 0.8 are considered as a interaction by the user, triggering the send of the command of action. In the test execution of the Figure 6 was intended the execution of

twenty-nine interactions, with different functions icons, which all of them been recognized correctly.

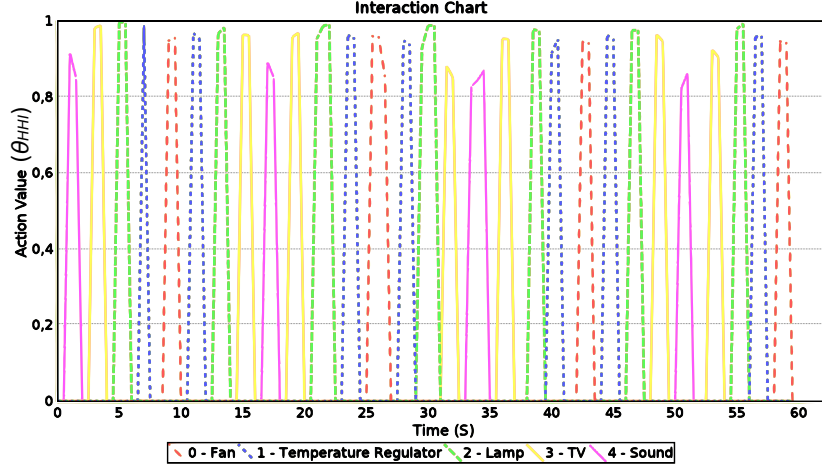


Fig. 6. Graphic of interaction in relation of the function action to the respective value of θ_{HHI}

The tests presented consistent results in both settings of light and scenarios, reaching high accuracy values. It is remarkable that none action have been triggered instead of another, corroborating the efficiency of the computer vision model adopted. Also, it is noteworthy that unsuccessful interactions are defined as the no recognized user interaction by the model, and not a interaction action executed arbitrarily, the last case being a more problematic situation.

Lastly, the interactions actions have been detected with comfortable margins, with values of θ_{HHI} close of one, for the actions actually desired, and zero for the others, demonstrating robustness of the model.

5 Conclusions

In this work, a model of interaction based on artificial intelligence techniques and computer vision was proposed for smart homes without use of sensors or controllers. The proposed model was able to detected and recognized the icons of interaction correctly in a uncontrolled environment, with different distances from the camera used and the target icons, and different luminosity variances in the ambiance, rarely being necessary parameters changes or any type of calibration.

As presented in the previous section, the results exposed that interactions made by the user have been correctly identified in almost every action of interaction, which reinforces the confidence in the model been proposed.

Additionally, our model works in day and night conditions, only using a ordinary camera, dispensing special sensors or another controllers, with allows a

low cost, easy learning and more natural form of interaction only using gestures. These characteristics was possible by the use of artificial intelligence and computer vision techniques, directing the complexity and cognitive enforces to the model and not to the users.

Conclusively, the model demonstrated satisfactory results, concretely offering a new option to control a residence. As future works, improvements in the method for detect interactions can be investigated, applying more intelligent approaches. Furthermore, a segmentation technique, providing a lower computational cost, and better segments quality, could be developed.

Acknowledgments. The authors would like to thank the Brazilian National Research Council (CNPq/PIBIC/UFS/PVE47492016) for the financial support provided for this research.

References

1. Chan, Marie and Hariton, Cyril and Ringear, Patrick and Campo, Eric: Smart house automation system for the elderly and the disabled, Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century., IEEE International Conference on, vol. 2, 1586–1589, IEEE (1995)
2. LEE, Kwang Yeol; CHOI, Jae Weon. Remote-controlled home automation system via bluetooth home network. In: SICE 2003 Annual Conference, p. 2824-2829, IEEE (2003).
3. AHMIM, Abdelhakim et al. Design and implementation of a home automation system for smart grid applications. In: Consumer Electronics (ICCE), 2016 IEEE International Conference on, p. 538-539, IEEE (2016).
4. Valli, Alessandro. "Natural interaction white paper." Recovered from <http://www.naturalinteraction.org/images/whitepaper.pdf> (2007).
5. Aggarwal, J. K., and Q. Cai.: Human motion analysis: a review. Nonrigid and Articulated Motion Workshop. In: Proceedings. IEEE 10 (1997)
6. Erol, Ali, et al.: A review on vision-based full DOF hand motion estimation. Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference, IEEE (2005).
7. Bharambe, Apoorva, et al.: Automatic Hand gesture based remote control for home appliances. International Journal of Advanced Research in Computer Science and Software Engineering 5.2 (2015).
8. Jain, Anil K., and Chitra Dorai.: Practicing vision: Integration, evaluation and applications. Pattern Recognition 30.2 (1997): 183-196.
9. Sumanta Dey, et al.: Gesture controlled home automation. International Journal of Emerging Engineering Research and Technology (2015).
10. Pannase, Dhanashree.: To analyze hand gesture recognition for electronic device control. International Journal 2.1 (2014).
11. Itti, Laurent, et al.: A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on pattern analysis and machine intelligence 20.11 (1998), 1254-1259.
12. Klava, B. Segmentação interativa de imagens via transformação watershed. Diss. Masters thesis, Instituto de Matemática e Estatística-Universidade de São Paulo, 2009.